IJCSE International Journal of Computer Sciences and Engineering Open Access
Research Paper Vol.-7, Special Issue, 6, March 2019 E-ISSN: 2347-2693

A Study on Scalable and automated Causal Relationship Discovery Strategies In Data Analytics

S. Sajida[1*], M. Padmavathamma[2]

[1,2]Dept of Comp. Science, SVU CM&CS College, S V University, Tirupati, India

Corresponding Author: sajidareddys@gmail.com, Tel.: 07780640950

DOI: https://doi.org/10.26438/ijcse/v7si6.915 | Available online at: www.ijcseonline.org

Abstract**—** Causation is one of the crucial relationships among the related variables that provide good stuff for data analytics. A causal relationship among a set of events exists when one or group of event/s is the result of the occurrence of the other event or a set of events. The primary objective of any research in data analytics or scientific analysis is to identify the level to which a relation exists among the subjective variables. Causal research can facilitate business environment to quantify the effect of present business practices on future production levels to aid in the business planning process. The process of discovering causal relationships among variables have multitude application areas like critical care services in medicine, advertising, bioinformatics, road safety ,share markets, and too more to be included. The present work targeted to study the existing methods of causal relationship discovery. The study also tried to propose the automated and straight forward causal relationship discovery methods which are scalable.

Keywords— *Decision tree*, causal relationship,bayesian netorks,Structural Equation models,CDT

I. Introduction

Causal inference, carrying out cause from effect, is possibly one of the most important problems to be dealt with. Designed experimental support is needed to make absolute statements about cause and effect. The set up these experiments is very expensive, and sometimes impossible.
Naturally people recognize causal relationships in their life journey. One may infer the cause of an event based on observation. Hard work leads to good results. Healthy intake of food causes better health. Sometimes the same instance may be a cause and an effect as well. Causal relationships help policymakers, practitioners, and scientists by providing them the cause and effect pair estimations. Among the sets of possible cause and effect relation pairs most of the candidates were neither feasible nor desirable and a few only are credible. The present and past three decades computer science, research was attracted by causal discovery methods from observational data. Currently Bayesian network techniques contributing the core of the methodologies for causal discovery in computer science. Structural equation models(SEMs) are also following the way.

The present paper organization started with the introduction to causal inference. In the next section the existing causal inference approaches were discussed. The main objective of this paper is to review the current approaches of causal relation discovery. This review was presented in the followed section. The summery of the review along with the need and scope of the future work was highlighted next.

1. Existing Causal Inference Approaches

Causal relationships were mostly studied using statistical techniques. Bayesian theorem was the foundation to analyze and predict a cause from the observed sets of effects under study. This approach can give probabilistic predictions. In critical decision contexts probabilistic reasoning alone is not sufficient to rely upon. In general, it may not be possible to find a reliable true causal model for the given inputs. Sometimes it may be very difficult to find performance of algorithms of causal models. Frequently, a causal model is described in two parts- the first part is called a statistical model and the second part is called a causal graph that describes the causal relations between variables. Among the causal models, two are commonly used in many situations. The first one is causal Bayesian networks and the second model is structural equation models (SEMs).

Bayesian belief networks: A Bayesian Belief Network deals with joint conditional probability distributions. Other names for this approach are Belief Networks, Bayesian Networks, or Probabilistic Networks.
Definition: A *Bayesian network* is a pair <G,P>, where G is a directed acyclic graph (DAG) whose vertices are random variables, and P is a density such that each variable V in G is

independent of variables that are neither descendants nor parents of *V* in *G*, conditional on the parents of *V* in *G*. In this case *P* is said to satisfy the *local directed Markov condition* for *G*.

The definition of class conditions between subsets of variables are the inputs to a Bayesian Network. The learning is performed using graphical model of causal relationships .Trained Bayesian Network provides a model for classification. Bayesian networks provide a graphical representation of conditional independence among a set of variables. A node represents a variable and, a directed edge between two nodes in a Bayesian network represents a causal relationship between the two variables [17].Many improved and current causality based methods use causal Bayesian networks (CBNs) either directly or indirectly to generate directed acyclic graph (DAG) in order to represent conditional independence between the variables. Many researchers in the area of cause/effect relations have observed that computational cost of finding complete or local causal graphical structures is very high. Hence, many constraint based algorithms were developed and implemented successfully for finding causal structures and obtained good results. Note that constraint based methods use observational data for finding conditional independence between variables based on user supplied threshold values. In the case of complex causal based systems it is highly difficult to set appropriate minimum and maximum threshold values. Experience has shown that error rate of false negatives increases when minimum support decreases and as a result of this redundancy increases in the system and making the system inefficient. Similarly the error rate of false positives increases when minimum support increases which results no causal rules will be found.

For the past three decades, many algorithms have been developed and reported for learning Bayesian networks from data. Inference in Bayesian networks can be seen in two ways .The first one is exact inference in which the conditional probability distribution is analytically computed over the selected variables. Sometimes the exact inference is hard to get. In such cases approximation techniques are used based on sampling. From Bayesian networks we can get the answers for the questions like "What is the conditional probability?","What is the value of a subjective variable at which the posterior probability is maximum?" and "What is the whole probability distribution?".The chain rule of Bayesian networks provides the joint probability as a product over the nodes of the probability of each node's value given the values of its parents.). In Causal Bayesian networks a density for a variable is specified as a function of the values of its causes where as in Structural equation models the value of a variable is specified as a function of the values of its causes. However, up to now it is only feasible to learn a Bayesian network with limited size of variables. Therefore, in practice the scalability of this approach is a challenge.

Structural equation models: Structured equations are widely used in social sciences. The search in a causal model is divided into two steps: searching for a desired graph and then estimating the free parameters from the desired causal graph and the sample data. Standard statistical methods are used for estimation the free parameters. SEM method takes three inputs and produces three outputs. The inputs are i1) A set of qualitative causal assumptions. i2) A set of queries concerning causal and counterfactual relationships among variables of interest. i3) A set of experimental or non-experimental data, governed by a joint probability distribution. The outputs are o1) A set of statements that are the logical implications of the first input. o2) A set of data-based claims concerning the magnitudes and o3) A list of testable statistical implications of i1.Structural equation models (SEMs) become the most popular approach to causal analysis in the area of social sciences. The capability of SEMs to support causal inference has got the significance. The trustworthiness of this model is still a debate.

2. The Role of Decision Tree in Causal Inference

Discovery of causal relationship is a type of supervised learning with a label is fixed for a target /outcome. In such cases classification methods are capable of finding the signals of causality. In addition classification methods are fast and can be used as substitutes for causality discovery techniques/methods of data science .One of the prevalent method of classification is Decision tree, which have been in use in many areas including social, business and medical data analyses. Causal inference and decision analysis are two areas of statistics. The overlap between the two areas is little. Decision tree methodology is a well known data mining method for setting up classification systems based on multiple covariates and for developing prediction algorithms for a target label. This model classifies a population into subdivisions that construct a tree with the whole data at root along with internal nodes, and leaf nodes representing the path and target class. The algorithm is able to deal with large, complicated datasets without imposing a complicated parametric structure and thus it is non-parametric. The model is developed with training part of a large data set and validation of datasets is done by test datasets.

Causal inference needs continuous reassessment whereas decision tree analysis is more of an inference down a probability model. At a fundamental level of thinking and application causal inference and decision analysis are connected .At advanced level of application both went on separately unless the some connection criteria is defined. The two approaches allow manipulation and led to potential outcomes. In causal inference, the "causal effect is the difference among what would happen under the n number of specified treatments. In the case of decision analysis the concern was on what would happen at one of the decisions taken. The research in these two areas may or may not closer. As classification methods are not intended for causal discovery in mind there is a possibility that classification methods may find false causal signals in data and the true causal signals may be missed. As classification method ignore the effects of other variables on the class label or outcome variable at the time of relationship examining between a variable and the class label, it may leads to false discoveries of causal relationships. To get the true causalities it is needed to work up on the effects of all the variables in the subject of classification study. Therefore it is a fundamental challenge of present research to study the casual inference approaches and develop scalable and compatible tools to infer the causalities in current data.

II. RELATED WORK

Yeying Zhu, etal [20] studied a casual inference problem with a continuous treatment variable based on propensity scores. The authors defined Propensity score as the conditional density of the treatment level given covariates. Propensity scores were used to estimate the weights of inverse probability. A boosting algorithm was suggested to estimate the mean function of the treatment given covariates. Number of trees to be generated determines the trade-off between bias and variance of the causal estimator. Therefore this is a boosting and tuning parameter. A criterion called average absolute correlation coefficient (AACC) was proposed to determine the optimal number of trees. Authors pointed that the estimation of the generalized propensity scores is a much more challenging task. This work concluded that the proposed method performs better than the existing methods, especially when the function of the treatment given covariates is not linear.

Sara Magliacane, et al. [13] discussedabout Joint Causal Inference (JCI) by taking multiple data sets to learn both causal structure and outcomes interactively. They observed that JCI offers many advantages when compared with the many existing constraint based causal inferences for finding causal relationships from the pooled data of multiple data sets. The main advantage of this approach isthe ability of the method to learn intervention outcomes.Two more advantages include the improved accuracy in predicting causal relationships and the statistical power of testing independence. Authors have observed problem in JCI called faithfulness violation problem due to deterministic relations. An effective strategy was proposed for handling faithfulness violations and this strategy was implemented in a package called ACID, a determinism-tolerant extension of Ancestral Causal Inference (ACI).

Christopher D. Ittner [4]discussed about the need of qualitative and quantitative methods and the power of rich institutional knowledge in getting the convincing evidence regarding the causal links between accounting practices and organizational outcomes. The author argued that, empirical tests of causal relationships in non-experimental settings are susceptible to multiple threats to validity. The threats may include but not limited to interactions, correlated omitted variables and endo-geneity, nonlinearities, simultaneities, and measurement error. The discussion emphasized the urge of minimizing these threats in empirical tests. It was also pointed out that always a control is needed to against confounding influences that may affect the outcomes of interest in their regression or structural equations models.

Constantin F, et al [5] **introduced a new framework on** Generalized Local Learning (GLL), developed **and evaluated** local causal and Markov blanket induction algorithms. They have extended causal core methods after analyzing GLL framework thoroughly, studied how model prediction will improve with respect to increase in the sample size and also investigated sensitive details of algorithms to multiple statistical testing. They have extended many existing GLL features such as divide-and-conquer causal model strategies from local-to-global learning, causal graph concepts, and also causal model distributed versions of GLL algorithms.

Jiuyong Li, et al [9] proposed a framework for the concept of causal rules and developed an algorithm for mining causal rules in large training data sets. Fundamental goal of causal discovery is to find the cause-effect relationships between variables. For finding causal rules they used retrospective cohort studies which are based on the results of association rule mining. They conducted many real experiments using synthetic as well as real world training data sets and explained the effectiveness and efficiency of causal rules data mining. Causal rules are more advantageous than association rules. Authors proposed a new method and it can be used as a hopeful alternative for causal discovery in large and high dimensional data sets. They proposed a new approach which is fast, have the capability of finding a real cause consisting of multiple variables. They argued that causality relationship has been well studied and well utilized in various disciplines such as engineering, medicine, banking, epidemiology, zoology, biology, economics, chemistry, physics, social science, science as a basic, powerful, efficient and effective tool for analysis, design, decision making, reporting, explanation, and prediction.

In [14]matching methods are used under similar observations for finding cause effect relationships by comparing treated units with control units. Authors said that there exist different types of matching methods being used in various applications such as medicine, criminology, science, economics, education, social sciences, public policy analysis, scientific disciplines, statistics, machine learning, data mining, sociology, psychology, research, behavioral sciences and so on. Matching methods generally assume few methods on the data relationships between covariates, treatment, and output variables. Matching methods can be applied on different types of multidisciplinary applications. Matching methods mainly based on distance measures covariate values in the treatment and control groups.

Stephen L et al [18] Studied advanced features of matching estimators and listed practical limitations of matching estimators. The authors introduced matching methods by taking ideal scenarios based on stratification and weighting procedures and also discussed most important data analysis techniques by taking hypothetical examples. Authors argued that the matching techniques can also be used efficiently, correctly and effectively in order to strengthen the prosecution of actual causal questions in many branches of sociology. They said that fundamental goal of causal relationship analysis is to investigate the selected effects of a particular cause instead of searching all possible causes of a particular outcome with respective to relative effectives.

Jiuyong Li ,et al[10] said that causal relationships are generally found with designed experiments such as randomized controlled trails but these are costly to conduct. They also said that causal relationships can also be discovered with well designed observational studies by taking the help of domain expert's knowledge and also pointed out that this is a time consuming process. They observed that more advanced scalable and automated state-of-the-art techniques are needed for finding potential causal relationships between the variables and the outcome variable in the case of large data sets. Authors pointed out that classification methods may appear that they are good for finding causal relationships but in reality the classification methods may find false causal relationships and could miss true causal relationships. They studied that classification methods fail to take accounts of other variables while trying to establish causal relationships between the input variables and the outcome variable. Authors argued that classification methods are not designed for finding causal relationships and they proposed a new scalable, automated causal decision tree framework model based on special statistic based causal relationship framework for finding true causal relationships from the large data sets. The proposed new technique is also applicable for big data applications also.

In [15]authors studied in detail about overview of relations among the causal modelling methods and provided overview of four important causal models for health sciences research. These models are: graphical models, potential outcome models, sufficient component cause models and structural equations models. They discussed about logical connections among the different types of causal models and particularly strengths and weaknesses of each of the causal models. A graph is causal if every arrow represents the presence of an effect of the parent variable on the child variable. Graphical models can display broad qualitative assumptions about causal directions and independencies in a population. Note that potential outcome models are not inherently deterministic because of the reason that the potential outcomes may be parameters of probability distributions rather than directly observable events. Authors said that the graphical causal models saw an explosion of various theoretical developments during the years 1990s including clear elaboration of connections or links to many other useful methods for finding causal relationships using causal modelling techniques including both endogenous and exogenous. Causal relationship based methods generally leads to define better models that are logically better-quality than either potential-outcome models or graphical models.

Bollen K.A ,et al.[3]presented a set of myths are following the trustworthiness of the model include :(1) SEMs aim to establish causal relations from associations alone, (2) SEMs and regression are essentially equivalent, (3) no causation without manipulation, (4) SEMs are not equipped to handle nonlinear causal relationships, (5) a potential outcome framework is more principled than SEMs, (6) SEMs are not applicable to experiments with randomized treatments, (7) mediation analysis in SEMs is inherently no causal, and (8) SEMs do not test any major part of the theory against the data. The authors presented the facts that chase away these myths. It is clearly described about what SEMs can and cannot do. The authors pinpointed that the current capabilities of SEMs to formalize and implement causal inference tasks are indispensible. The potential of SEMs to do more was emphasized.

Peter Spirtes [17] has discussed problems of large number of variables, small sample sizes, usage of unmeasured causes and pointed out that these problems are occurring in many real time applications. The author also discussed all these causal relationship determining problems during application of graphical causal modeling algorithms. The author reminded many problems in the domain of causal modeling and some of the problems listed by him are – how to match causal models and search algorithms to causal problems, model selection, and prior knowledge, how to improve efficiency and efficacy of search algorithms, characterization of search algorithms, addition and deletion of simplifying algorithms. Author also discussed about the actual problems of causal inference, described several different kinds of causal models, discussed about the potential problems that are associated with search for causal models. Author mainly pointed out that why algorithms appropriate for finding good classification or prediction models in machine learning are not always appropriate for finding good causal models.

M.W. Birch [2] has discussed the point that how to test partial association between two random variables x and y in a three-way contingency table. He revealed an important point that for a long time zero partial association neither implies nor is implied by zero total association. Mantel-Haenszel test gave a refined version of Cochran's second test statistic which remains valid when the number of component tables is large and the individual frequencies are small, based on the hyper-geometric distribution. In this paper author has shown that the Mantel-Haenszel test is optimal for testing against alternatives in which the degree of partial association is constant and the test criterion has a chi-squared distribution with 1 degree of freedom on the null hypothesis. A number of measures have been proposed for finding quantitative measure of association. Mantel and Haenszel have given a straightforward application of their test to some lung-cancer/cigarette-smoking data set.

Swati Hira and P. S. Deshpande [19] have proposed a novel framework for extracting cause-effect relationships in large time series data sets containing socioeconomic indicators. They have extended the existing cause-effect relationships by introducing new multiple causal structures such as binary, transitive, many to one and cyclic. In order to obtain high reliable causal rules they have used temporal association and temporal odds ratio. Authors have used both synthetic and real-world data sets for performance computations. Proposed method is very useful for building quantitative models to analyze socioeconomic processes by generating potential cause-effect relationships. The proposed method is useful in various application domains such as research, business, social, science, economic, agriculture to generate strategic rules for effective decision making dynamically. The proposed method is also very useful for finding exact cause of fault in the case of large mechanical system which is monitored by various sensors during generation of time series data.

Li. j ,et al[12] have presented a PC-simple algorithm for exploration of local causal relationships with respect to output variable. Authors pointed out that probabilistic graphical model, Bayesian network, is the most important model for exploring causal relationships in data sets. Bayesian network completely represents joint probability of all the variables using directed acyclic graph (DAG). The PC- simplified algorithm is a commonly used simplified method for learning the structure of a causal Bayesian network (CBN).

C. F. Alfieri's, et al. [1] proposed a framework called Generalized Local Learning (GLL) for local causal relationship exploration with respect to the output variable. This framework obtains causal inferences from very large data sets in the form of cause/effect and Markov blankets. Authors pointed out that local causal relationship determination methods can be used for the purpose of scalable and accurate global causal graph learning.

Donald B. Rubin [6] said that the main aim of observational studies on very large data sets is to find the causal effects after applying new treatments subjected to the conditions. Author was applied propensity score method for finding causal relationships. Author said that very large databases are the potential sources for applying causality based questions and observed that standard statistical models are not able to handle such situations and propensity score methods are right tools for elicitation of the causal relationships. Author also pointed out that any causal related question must be first approached with propensity score method for find intensity of causal effects and then be applied with the selected method.

L. Frey ,et al.[7] presented a well designed framework with local causal relationship determination capability with respect to the targeted variable but it can be used for finding global causal relationships very easily. Working principle is local but applicability is global. Its main advantage is that it can be applied on very large data sets but with small sample sizes. This framework consists of many state-of-the-art tools for effective exploration of causal relationships.

P. Spirtes ,et al.[16] said that a set of assumptions and methods are needed for finding causal inference relationships in the data sets and it is possible to apply incomplete causal knowledge on many real time applications such as scientific, engineering, science, prediction, social, behavioral, planning etc,. Authors pointed out those experimental and observational studies may not always produce same inferences. They studied and experimented on many data sets for finding causal relationships using both structural and categorical models. Authors also said that the relationship between causality and probability can also be useful for finding clarifications in many topics of statistics.

Jiuyong Li,et al. [11] presented four important methods for finding causal relationships in many diversified applications such as economics, research, physics, business, biology, engineering, medicine, epidemiology, social sciences. Authors said that some methods use conditional independence and still other methods use association rule mining for discovering cause-effect relationships in the data sets. Authors realized that all of the four methods innovative and effective for identifying potential causal relationships with respect to the given output variable, and each method has its own advantages and disadvantages. Many potential and useful methods are designed and developed for finding causal relationships in many areas such as business, artificial intelligence (AI), banking sectors, machine learning, medicine, data mining, and biomedical research.

Z. Jin ,et al.[8] statedthat finding causal relationships in large databases of observational data is very difficult. Bayesian networks are predominantly used in this area for discovering causal relationships but only its main disadvantage is that Bayesian network learning is a NP-complete problem, as a result of this many constraint based algorithms have been designed and developed for effective discovery of causal relationships from large data sets. All these new methods are based on Bayesian learning either directly or indirectly and uses single cause variable in causal relationships exploration. Authors have proposed a new approach for finding causal relationships form the very large data sets without predefining any thresholds. They said that causal relationship is more powerful than associated relationship. Finding complete or local causal inferences using causal graphical models need very high computational cost and to overcome this problem constraint based algorithms were proposed.

Zhou JIN, et al.[21] proposed a non-graphical approach for finding top-k causal rules (TKCL) and an efficient algorithm was developed for mining top-k causal rules for finding causal relationships form the very large data sets without predefining any thresholds. They conducted experiments on both real and synthetic data sets and verified that the proposed algorithm is effective in finding top-k causal rules from the large data sets and the algorithm is scalable.

III. METHODOLOGY

This systematic study covered various methods and applications in the context of casual inference. Each method proposed in the literature have its own merits and limitations in spite of a specific context undertaken. There exist different types of causal models that are using in different disciplines. A graph is said to be causal if every arrow represents the presence of an effect of the parent variable on the child variable. A direct path of the causal graph represents a causal pathway. In general, causal models are categorized into Graphical causal models (causal diagrams), Potential outcome models (counterfactual models), Sufficient-component cause models and structured equation models (SEMs).

The review found that graphical causal models are popular in discovering causal relationships and particularly causal Bayesian networks (CBNs) are most suitable and have been in use for a long time associated with many real world applications for causal inference exploration. Structure Equation Models (SEMs) are also becoming vital overcoming the barrier of criticism about their applicability for casual inference. Causal discovery methods can be broadly divided into two categories. The first category of methods was based on causal Bayesian network and the second category comes under standard data mining techniques combined with statistical methods. The existing decision tree classification models were actually designed for classification and not for causation. The underlying reason for this limitation is that the classification process ignores the combined effect of all variables available. To find true causal relationships from the data set the decision trees needs to be enforced with some statistical methods and the approach of classification to be embedded with casual modeling Scalable causal decision tree frameworks for discovering interpretable and context specific causal relationships from the large data sets needs to be proposed and developed. The purpose of causal structure discovery is to find causal relationships in the data sets. No standard causal relationship framework is available for finding causal inferences which leads to finding causal relations in observational data is a challenging task. Main advantage of constraint based approaches in determining causal relationships is that they do not need complete graphical structure and some of such models are – LCD, GLL, CCC,and CCU. Constraint based methods generally produce fixed partial outputs in the form of directed acyclic graph (DAG) and it shows that these methods use Bayesian learning directly or indirectly in some means.

The need for advanced scalable approaches

Though the availability and the ability of existing methods for casual inference are encouraging, more advanced methods are needed to cope with the current data analysis needs. The availability of data in variety of forms, at high volumes is becoming a key challenge. From this data with large set of attributes, the selection of right set of attributes (parameters) for causation process is again a big task. Existing statistical, data mining and machine learning (ML) approaches to estimation, model selection and robustness do not directly apply to the problem of estimating causal parameters.

IV. RESULTS AND DISCUSSION

The need for advanced scalable approaches

Though the availability and the ability of existing methods for casual inference are encouraging, more advanced methods are needed to cope with the current data analysis needs. The availability of data in variety of forms, at high volumes is becoming a key challenge. From this data with large set of attributes, the selection of right set of attributes (parameters) for causation process is again a big task. Existing statistical, data mining and machine learning (ML) approaches to estimation, model selection and robustness do not directly apply to the problem of estimating causal parameters. There is a need to develop robustness measures for finding causal parameters. For causal questions, one must know what would happen if an antecedent changes its policy.Conditional probability trees (CPT-trees) and causal explanation trees works on known causal relationships at the beginning. Nowadays machine learning methods particularly decision tree methods are employed for discovering causal relationships in many different types of sub-domains. There is a need to develop the tree models with the capability of casual relationship discovery without presumptions.

V. Conclusion and Future Scope

The present review has undertaken the challenge of casual relationship discovery in large data sets. A good number of research articles were reviewed to understand the existing casual relationship discovery approaches in practice. The review identified the commonalities, differences among the approaches and found the limitations with respect to the current trends and needs of data analytics. The Causal Bayesian Networks and Structural Equation Models were identified as the most promising approaches. But the scalability, automation and robustness of these models were the limiting factors. Decision tree was also a trustable mean for casual inference when applied on context specific data with limited effecting variables. The review identified the need of hybrid approaches that combine the power of classification models, statistical models and data mining techniques. This study offered that, a decision tree with causal relationship discovery as the prime target will provide the scalable and reliable model.

REFERENCES

[1]. Alfieri's C. F., et al. [C. F. Aliferis, A. Statnikov, I. Tsamardinos, S. Mani, and X. D.Koutsoukos, "Local causal and Markov blanket induction for causal discovery and feature selection for classification part I: Algorithms and empirical evaluation", Journal of Machine Learning. Res., vol. 11, pp. 171–234, 2010.

[2]. Birch M.W." The Detection of Partial Association", Journal of Royal Statistical Society Aeries B (Methodological), Vol. 26. No. 2 (1964), pp. 313-324.

[3]. Bollen K.A., Pearl J. "Eight Myths About Causality and Structural Equation Models. In: Morgan S." (eds), Handbook of Causal Analysis for Social Research. Springer, Dordrecht (2013)

[4]. Christopher D. Ittner, "Strengthening causal inferences in positivist field studies", Accounting, Organizations and Society 39 (2014) 545–549.

[5]. Constantin F. Aliferis, Alexander Statnikov, Ioannis Tsamardinos, Subramani Mani, Xenofon D. Koutsoukos, "Local Causal and Markov Blanket Induction for Causal Discovery and Feature Selection for Classification", Journal of Machine Learning Research 11 (2010) 171-234.

[6]. Donald B. Rubin [Donald B. Rubin, "Estimating Causal Effects from Large Data Sets Using Propensity Scores" 15 October 1997 | Volume 127 **Issue 8 Part 2 | Pages** 757-763, Annals of Internal Medicine, American College of Physicians]

[7]. Frey L., D. Fisher, I. Tsamardinos, C. Aliferis, and A. Statnikov, "Identifying Markov blankets with decision tree induction," in Proc. 3rd IEEE Int. Conf. Data Mining, Nov. 2003, pp. 59–66.]

[8]. Jin Z., J. Li, L. Liu, T. D. Le, B. Sun, and R. Wang, "Discovery of causal rules using partial association," in Proc. IEEE 12th Int. Conf. Data Mining, Dec. 2012, pp. 309–318]

[9]. Jiuyong Li, et al [Jiuyong Li, Thuc Duy Le, Lin Liu, Jixue Liu, Zhou Jin, Bingyu Sun, and Saisai Ma", From Observational Studies to Causal Rule Mining". ACM Trans. Intell. Syst. Technol.2015

[10]. Jiuyong Li, Saisai Ma, Thuc Duy Le, Lin Liu and Jixue Liu "Causal Decision Trees," arXiv: 1508.03812v1 [cs.AI] 16 Aug 2015]

[11]. *Jiuyong Li*, *Lin Liu*, *Thuc* Duy *Le*. *Jiuyong Li • Lin Liu • Thuc* Duy *Le, "Practical Approaches to Causal Relationship Exploration" Jiuyong Li* School of Information Technology and Mathematical Sciences University.2015]

[12]. Li. j, Liu. L, Le. T., "Practical approaches to causal relationship exploration"2015. X. 80 p. 55 illu., softcover, ISBN:978-3-319-14432-0, http://www.springer.com/978-3-319-14432-0]

[13]. Magliacane Sara, Tom Claassen, Joris M. Mooij, "Joint Causal Inference from Observational and Experimental Datasets", Journal of Machine Learning Research, March 2017.

[14]. S. L. Morgan and D. J. Harding, "Matching estimators of causal effects: Prospects and pitfalls in theory and practice," Sociological Methods Res., vol. 35, pp. 3–60, 2006.

[15]. Sander Greenland and Babette Brumback, "An overview of relations among causal modelling methods",*International Journal of Epidemiology*, Volume 31, Issue 5, 1 October 2002, Pages 1030–1037.

[16]. Spirtes P., C. C. Glymour, and R. Scheines, Causation, Predication, and Search, 2nd ed. Cambridge, MA, USA: MIT Press, 2000.

[17]. Spirtes Peter. "Introduction to Causal Inference, " Journal of Machine Learning Research 11 (2010) 1643-1662.

[18]. Stephen L. Morgan, David J. Harding, "Matching Estimators of Causal Effects", Sociological Methods& Research Volume 35 Number 1August 2006 3-60 _ 2006 Sage Publications

[19]. Swati Hira and P. S. Deshpande ,"Mining precise cause and effect rules in large time series data of socio - economic indicators," *SpringerPlus (2016) 5:1625.*

[20]. Yeying Zhu, Donna L. Coffman and Debashis Ghosh, "A Boosting Algorithm for Estimating Generalized Propensity Scores with Continuous Treatments", J. Causal Infer. 2015; 3(1): 25–40.