2018 5th International Conference on Electrical & Electronics Engineering and Computer Science (ICEEECS 2018)

# Research on Structure Learning Theory and Algorithm Based on Probabilistic Graphical Model

**Li Feng**

Department of Mathematics and Statistics, Xidian University, Xi'an, Shaanxi, China, 710071

**Keywords:** probabilistic graphical model; Bayesian network; Markov network

**Abstract:** Probabilistic graphical model can effectively deal with uncertainty reasoning, and learning the probabilistic graphical model from the sample data is an important problem in practical application. The representation of the probabilistic graphical model consists of two parts, parameters and structure. The learning algorithm is also divided into parameter learning and structure learning. In this paper, the structure learning algorithms based on probabilistic graphical model network are introduced in detail. The structural learning algorithms are also summarized based on the differences in the characteristics of structural learning algorithms, and the algorithm of Markov network structure learning is summarized. Finally, this paper points out the open problems of the probabilistic graphical model learning and its further research directions.

## 1. Introduction

Probabilistic graphical model can represent the interrelationships between variables succinctly and effectively, and provide powerful tools for uncertain reasoning systems. In recent years, they have become hot areas for artificial intelligence and machine learning. At present, probabilistic graphical model have been used in image analysis, and applied in many fields such as biomedicine and computer science [1]. Before using probabilistic graphical model for indefinite reasoning, it is necessary to construct probabilistic graphical model based on domain knowledge. In the past, the construction of probabilistic graphical model is usually constructed manually by domain experts using professional knowledge, but the process is time consuming and the implementation process is complex. In today's information age, learning the probabilistic graphical model from the sample data has become the main means of construction.

## 2. Probabilistic Graphical Model

Probabilistic graphical model combine graph theory with probability theory to describe multivariate statistical relationships compactly. Probabilistic graphical models have multiple representation types, such as Bayesian networks, Markov networks, Chain graphs, transient models and Probabilistic relational model. Although these types of representations vary, but the main idea is to use conditional independence assumptions to factorize the joint probability distribution, simplify the representation and reasoning calculation process [2]. This paper mainly studies the learning algorithms of Bayesian network and Markov network.

### 2.1 Bayesian network.

BN can represent the random variable joint probability distribution set $\boldsymbol{X} = \{X_1, \cdots, X_n\}$, which consists of two parts: network topology and parameters. The structure of BN is Directed acyclic graph, where the node represents a variable, and the directed edge represents a conditional dependency between variables. The parameter of BN is the conditional probability distribution of the random variable of the node, that is, the conditional probability distribution of the variable when the parent node is known. According to the implicit independence assumption in the structure: When knowing the parent node, $X_i$ is independent of its non-children condition, then the joint probability distribution can be decomposed into the product of multiple conditional probability distributions:

 

$$P(X_1,\cdots,X_n)=\prod_i P(X_i|Pa_{X_i}) \quad (1)$$

Where, $Pa_{X_i}$ represents the parent node of the variable $X_i$.

Due to BN's strong reasoning ability, BN has been successfully applied in many fields such as medical diagnosis, clinical decision making, bioinformatics, forensic medicine, speech recognition, risk analysis, and reliability analysis.

### 2.2 Markov network.

MN is a type of probabilistic graphical model representing the symmetry influence relationship between random variables. It is also composed of topological structure and parameters. The structure of MN is undirected graph, nodes represent random variables, and no-directed edges represent dependencies between variables. The variables in the MN structure can be divided into multiple clusters, and then the joint probability distribution can be decomposed into the product of the factors of each cluster:

$$P(X_1,\cdots,X_n)=\frac{1}{Z}\prod_{k=1}^{m}\emptyset_k\,(C_k) \quad (2)$$

Where $m$ is the number of clusters, $\varphi_k$ is the factor of the $k$ cluster, and $C_k$ is the set of random variables for the $k$ cluster,

$$Z=\sum_{X_1,\cdots,X_n}\prod_{k=1}^{m}\emptyset_k\,(C_k)$$

is a partition function, in the learning task, the MN is usually represented by a log-linear model, that is, each factor is expressed as an exponentially weighted sum of the eigenfunctions of the variable set:

$$P(X_1,\cdots,X_n)=\frac{1}{Z}\exp(\sum_i \omega_i f_i(C_i)) \quad (3)$$

The characteristic function $f_i(C_i)$ is any real-valued function of the variable set $C_i$, $\omega_i$ is the weight of $f_i(C_i)$. MN is most commonly used in computer vision and image processing to model the relationship between pixels.

## 3. Structure Learning Algorithm

### 3.1 Constraint-based learning algorithm.

CB learning algorithms typically use CI tests or mutual information to identify dependencies and independent relationships among variables, and then establish networks that satisfy these relationships [2]. However, the performance of the CB algorithm depends on the number of CI tests and the size of the constraint set. The larger the number of higher-order CI tests is, the lower the accuracy of the algorithm is. Therefore, the CB algorithm is generally suitable for sparse Bayesian networks. Furthermore, the structure learning process is highly sensitive to test errors when a certain CI test is performed. When mistakes occur, it will directly mislead the follow-up inspection results.

### 3.2 Score-based learning algorithm.

The SS learning algorithm treats the structural learning problem as a model selection problem and consists of a scoring function and a search algorithm. The scoring function is used to evaluate the fitting degree of the candidate structure and the data. And the better the fitting, the higher the score. The search algorithm is candidate structures are spatially searched for the highest scoring structure. However, since the candidate structure space size increases exponentially with the number of nodes, the search task is NP-hard. The search space is generally divided into three types: the space formed by the DAG, the space formed by the DAG equivalence classes, and the space formed by the sequence of variables [3]. Most search algorithms are based on DAG space development. When the search is in the DAG equivalence class space or variable sequence space, structure of the learning process will be faster.

Cooper and Herskovits first proposed a Bayesian scoring function for BN structure learning, called the *CH* score or the *K2* scoring function. When the structure S and the data set D are known, the *K2* scoring function is:

$$P(S,D) = P(S)\prod_{i=1}^{n}\prod_{j=1}^{q_i}\frac{(r_i-1)}{(N_{ij}+r_i-1)}\prod_{k=1}^{r_i}N_{ij}! \quad (4)$$

Where $n$ is the number of variables, $q_i$ is the number of parents of variable $i$, $r_i$ is the number of values of variable $i$, $N_{ijk}$ is the number of times that variable $i$ takes $k$ when parent is variable $j$, $N_{ij} = \sum_{k=1}^{r_i} N_{ijk}$

*P(S)* is the prior probability distribution of the structure *S*, and the other components in the equation are the likelihood functions of the structure when the data is known.

Most of the SS algorithms search in the DAG space. Even if the number of the largest parent node of each node is constrained, searching for the optimal structure in such a large space is still NP-hard. Due to the difficulty of the search task, many Heuristic search algorithms, such as greedy search, genetic algorithm, evolutionary programming, simulated annealing, particle swarm optimization and ant colony optimization Etc. Table 2 shows typical SS algorithms and lists their scoring functions, search algorithms, and search space.

Table 2 Classic scoring and searching learning algorithms

| Methods | Scoring functions | Searching procedure | Searching space |
|---|---|---|---|
| K2 | K2 | Greedy search | DAG |
| K3 | MDL | Greedy search | DAG |
| K2GA | K2 | Genetic algorithm | Ordering |
| HEA | CI test and MDL | Evolutionary rogramming | DAG |
| ChainACO | K2 | Ant colony optimization | Ordering |

### 3.3 MN structure learning algorithm.

The structure learning of MN can also use CB algorithm and SS algorithm. However, since the constraint assumption of CB algorithm in MN is rather harsh, it is usually not realized in practical applications, so its algorithm research is rarely. The most commonly used MN structure learning algorithm is SS learning algorithm. Since log-linear models are usually used in structural learning of MNs, for a pairwise MN, the weights corresponding to log-linear models between two nodes are not zero. Therefore, the structural learning problem can be regarded as a feature induction problem, and the corresponding weights can be learned. Based on the non-zero weights, the structure of the MN can be obtained. Therefore, the structural learning task of the MN is to find a high probability region in the sample space.

## 4. Research Trends of Probabilistic Graphical Model Learning

Although the learning algorithm of probabilistic graphical model has made important progress in many aspects, there are still many problems to be solved.

In the structural learning SS algorithm, most of the scoring functions are sensitive to parameters, or the optimal solution cannot be obtained in the case of a limited sample size, which limits the accuracy of structural learning [3]. Future research should further develop performance. A good scoring function improves the accuracy of learning.

The computational cost of the structural learning process is high, but in the real world it often contains a large number of variables, making it difficult to achieve structural learning. In addition to the parallel learning strategy, the structure learning process should be accelerated from the search algorithm in the future, and the intermediate calculation process should be optimized.

## 5. Summary

This paper systematically reviews the research progress of learning algorithms for BN and MN. This paper introduces the algorithms of structural learning in detail. Structural learning is the main part of learning tasks, and its algorithms can be roughly divided into constraints based learning algorithms, scoring search based learning algorithms, hybrid learning algorithms, dynamic programming structure learning, and model average structure learning. Compared with BN learning, MN research is not yet mature enough. This paper also studies the structure of MN. And the algorithm is briefly introduced. Finally, the new challenges of the probabilistic graphical model learning and its research trends are pointed out. The probabilistic graphical model is a powerful tool for uncertainty reasoning and is increasingly important in the field of machine learning. The probabilistic graphical model should be combining with the new theory and technology emerging in the field of machine learning, the theory and technology of its own system should be continuously improved so as to be better applied in practical fields.

## Acknowledgement

The National Natural Science Foundation of China,Learning and Reasoning Prediction of Deep Probabilistic Graphical Model,No.61573266

## References

[1] Koller D, Firedman N. Probabilistic Graphical Models: Principles and Techniques[J]. Cambridge: The MIT Press, 2009, p.32

[2] Wainsright M, Jordan M. Graphical Models, Exponential Families, and Variational Inference[J]. Foundations and Trends in Machine Learning, 2008, 1(1-2), p.305

[3] Korb K B, Nicholson A E. Bayesian Artificial Intelligence[M]. Florida: CRC Press, 2010, p.46