`Analysis of Probabilistic Graphocal Model Based on Markov Random Field and Conditional Probability

Genwang Zhou[1, a*], Jian Zhou[2, 3 b] and Saihan Wang[4, c]

[1]No. 60 Middle School of Lanzhou, Lanzhou 730060, China

[2]School of computer science, Dalian University of Technology, Dalian 116024, China

[3]Key Laboratory of China's Ethnic Languages and Information Technology, Northwest Minzu University, Ministry of Education, Lanzhou 730030, China

[4]School of Automation, Beijing University of Posts and Telecommunications, Beijing 100876, China

[a] zhougenwang79@163.com, [b]zhoujian@mail.dlut.edu.cn, [c]leonyi@bupt.edu.cn

[*]corresponding author

Keywords: Probabilistic Graphocal Model; Markov Random Field; Experiment Analysis

Abstract: This paper discusses the Probabilistic Graphocal Model (PGM) based on Markov random field and conditional probability. Through extensive verifications, it can be found that a random variable of the Markov random field is only related to its adjacent random variables, and is independent of random variables that are not adjacent. Furthermore, we analyze the potential function of Markov random field, Markov network connections are often expressed as a log-linear model. Finally, we also focus how the model to represent the uncertainty in the real world and the relationship between the various quantities. It is clear that the joint probability of the transition probability and the performance probability. The statistical probability is the conditional probability. Because it is only normalized locally, it is easy to fall into the local optimum.

1. Introduction

Probabilistic Graphocal Model (PGM) is a major research direction in the field of control AI, which plays a very significant role in the process of data analysis [1-4]. It is worth remarking that uses graphs to indicate the probability independent of certain variables. Supposing that the visualization knowledge of agent-based probability theory and graph network theory, the graph is used to represent the joint probability distribution of variables related to the model [5-6]. Developed by Turing Award winner Pearl. Probabilistic graph model theory is divided into probability graph model representation theory, probability graph model inference theory nodes and probability visualization graph modeling algorithm theory [7]. Over the past decades, the basic Graphical Model can be broadly divided into two categories: the Bayesian Network and the Markov Random Field [8-10]. This structural difference leads to a series of subtle differences in their modeling and inference. In particular, every node in a special Bayesian visualization network would be corresponds to the major stability and probability distribution for a random probability distribution modeling [11], thus the overall JIF distribution could be heavily assigned into the information of the each random nodes such as agent node and JIF node [12]. Of course, for a random Markov fields, since there exist explicit Gauss relationship between different variables, their thus joint random probability is always be represented as the information of a series of algorithm functions. Especially, the max-degree node of these visualization is not always equal to zero, so it must be normalized to form a valid probability random hub distribution.

The probability graph visualization modeling algorithm can be divided into the parameter learning and image registration. According to the probability graph model, the parameter learning and structure learning algorithms are divided into probabilistic networks [13-14], and are classified and determined into the learning algorithms according to the completeness of the data set [15]. Different, structural learning algorithms are summarized into constraint-based learning,

DOI: 10.38007/Proceedings.0000027 ISBN: 978-1-80052-000-4

scoring-based learning, mixed learning, dynamic programming structure learning, model mean network structure learning, and structural learning of positive data sets [16]. Of course, structural learning is still a very challenging direction in machine learning. Structural learning has no fixed form, and different researchers often take different approaches [17-18]. For example, a very important issue in structural learning is how to discover the internal relationships between variables. For this problem, people have proposed a variety of different methods: Another important development direction is non-parameterization. Unlike processed traditional agent-based methods, nonparametric approaches are a more flexible route to a nonparametric model (for example the number and value of network nodes) can change as the data changes. A typical nonparametric model is a hybrid model based on the Dirichlet Process. This model introduces the Dirichlet process as a prior distribution of different parameters, assigning for any number of components in the hybrid. This fundamentally overcomes one of the difficulties in the traditional AI modeling, which can be determined by the information of components. Over the past decades, nonparametric models can be used to research feature visualization learning. Therefore, in this regard, a more proper job is to attempt an indefinite number of features based on the Hierarchical Beta Process (HBP). Unlike optimization-based methods, the Monte Carlo method collects samples by running a random simulation of the probability model and then estimates the statistical properties (eg. mean) of the variables from the collected samples. The sampling method has three important advantages. First, it provides a method with a rigorous mathematical foundation to approximate the integrals that often occur in probability calculations (the complexity of the integral calculation grows geometrically as the spatial dimension increases). Second, the sampling process ultimately yields a sample set of the entire joint distribution, not just an optimal estimate of certain parameters or variable values. This sample set provides a more comprehensive depiction of the entire distribution. For example, you can calculate the correlation coefficient of any two variables. Third, its asymptotic behavior can usually be rigorously proven. For complex models, the solution obtained by variational inference or belief propagation is generally not guaranteed to be the global optimal solution to the problem. In most cases, it is not even possible to know how far it is from the optimal solution. If you use sampling, as long as the time is long enough, you can arbitrarily approximate the true distribution. Moreover, the complexity of the sampling process is often easier to obtain theoretical guarantees.

The Monte Carlo method itself is also a very important branch of modern statistics. Research on it has been very active over the past few decades. Common sampling methods in the field of machine learning include Gibbs Sampling, Metropolis-Hasting Sampling (M-H), Import Sampling, Slice Sampling, and Hamiltonian Monte Carlo. Among them, Gibbs Sampling is usually regarded as a special case of M-H because it can be interpreted in the M-H method - although their initial motivation is different. In th next section, we will analyze the applications of Probabilistic Graphocal Model and Monte Carlo simulation.

2. Model Formulation

To the best of our knowledge, this optimization problem may not be easily solvable. To this end, we first consider mapping the samples to the unit hypersphere, and then keeping the smallest sum of the samples and principal component orthogonal projection. The density distribution information for all samples can be reflected by the sum of these distances. The optimization problem thus can be transformed as follows:

$$\begin{aligned} &\min_{u} \sum_{x_i \in X} d\left(z_i, u\right) \\ &s.t. u^T u = 1 \end{aligned} \tag{1}$$

where $z_i = \frac{x_i - \overline{x}}{\left\| x_i - \overline{x} \right\|_2^2}$ $\left(d\left(z_i, u\right) = \frac{d\left(z_i, u\right)}{1} = \sin^2 \alpha\left(\left\langle x_i, u \right\rangle\right)\right)$.

More intuitively, above optimization problem can be turned as

$$\min_{u} \sum_{x_i \in X} \sin^2 \alpha\left(\langle x_i, u\rangle\right) \quad (2)$$
$$s.t. u^T u = 1$$

where $\sin^2 \alpha_i = \frac{\left\|\hat{x}_i - UU^T \hat{x}_i\right\|_2^2}{\left\|\hat{x}_i\right\|_2^2}$, $\sin \alpha_i = 0$ denotes that x_i in the density direction and coincides with the direction of principal component, while $\sin \alpha_i = 1$ shows that x_i in the density direction and is orthogonal to the direction of principal component. The visualization of PGM is show in Fig. 1.

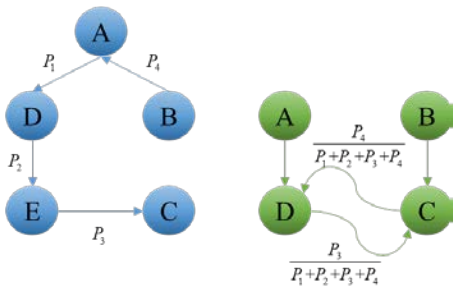

Figure 1. Visualization of PG Model

3. Analysis of Markov Random Visualization

It is clear that the most issues are strongly influenced by the uncertain factors such as communication and information visualization etc. That is, information visualization plays a very significant role in the process of calculation of JIF. Markov property: It refers to the distribution of a random variable sequence in chronological order. The distribution characteristics at the *N*+1th time are independent of the value of the random variable before *N*. In order to tackle the information for an analogy. Therefore, it can be found that the different weathers are Markov, so one speculate that variable JIF is only related to the probability of information visualization, and has nothing to do with the weather before and before the day before yesterday. Other laws of transmission such as infectious diseases and rumors are Markov's, such as illustrated in Fig. 2. The Markov nature is added on the basis of the random field to obtain the Markov random field. Mapping the Markov random field to the undirected graph, the nodes in the undirected graph are all related to a random assigned value, and the edges connecting the nodes represent the relationship between the random variables related to the two nodes, so The Markov random field actually expresses some factors that must be considered between random variables, while others may not be considered.

A random variable of the Markov random field is only related to its adjacent random variables, and is independent of random variables that are not adjacent. If random field $X = \{X_S, s \in S\}$ meets the following conditions:

$$\begin{cases} P = \{X = x\} > 0, \forall x \in A \\ P = \{X_s = x_s \mid X_r = x_r, r \neq s, \forall r \in \delta(s)\} \end{cases} \quad (3)$$

which can be called the modeling visualization of the Markov random field.

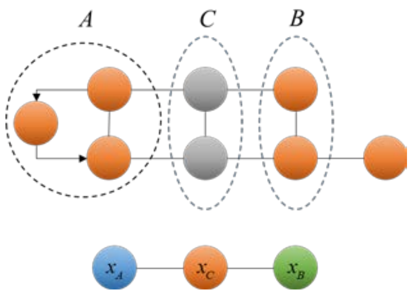

Figure2. A simple Markov Random Field

Therefore, Joint probability of Fig. 2 can be written as follows:

$$P(x_A, x_B, x_C) = \frac{1}{Z}\psi AC(x_A, x_C)\psi BC(x_B, x_C) \tag{4}$$

Based on the definition of conditional probability, the Eq. 4 can be rewritten as the following model.

$$\begin{aligned} P(x_A, x_B \mid x_C) &= \frac{P(x_A, x_B, x_C)}{P(x_C)} = \frac{P(x_A, x_B, x_C)}{\sum_{x'A}\sum_{x'B}(Px'_A, x'_B, x'_C)} \\ &= \frac{\psi AC(x_A, x_C)}{\sum_{x'_A}\psi AC(x'_A, x_C)} \cdot \frac{\psi BC(x_B, x_C)}{\sum_{x'_B}\psi BC(x'_B, x_C)} \end{aligned} \tag{5}$$

$$P(x_A, x_B \mid x_C) = \frac{P(x_A, x_C)}{P(x_C)} = \frac{\psi AC(x_A, x_C)}{\sum_{x'_A}\psi AC(x'_A, x_C)} \tag{6}$$

From Eq. (5) and Eq. (6), we can obtain follows:

$$P(x_A, x_B \mid x_C) = P(x_A \mid x_C) \cdot P(x_B \mid x_C) \tag{7}$$

Then x_A and x_B are independent when x_C is given.

Next, we would like to analyze the potential function of Markov random field. Clearly, the construction of the potential function is a key issue in the artificial potential field method. The potential function is a mathematical function of the physical vector potential or the scalar potential, also known as the harmonic function, which is the research topic of the mathematical potential theory.

For example, the variables of Fig. 2 are binary variables, if the potential function can be formulated as:

$$\psi AC(x_A, x_C) = \begin{cases} 1.5, if\ x_A = x_C \\ 0.1, otherwise \end{cases} \tag{8}$$

$$\psi BC(x_B, x_C) = \begin{cases} 0.2, if\ x_{AB} = x_C \\ 1.3, otherwise \end{cases} \tag{9}$$

It means that the model preference variables x_A and x_B have the same value, and x_B and x_C have different values. In this model, x_A is positively related to x_C, while x_B is negatively related to x_B. In fact, Markov network connections are often expressed as a log-linear model. By introducing feature functions Φ_k, which can be denoted as follows:

$$f_k = \exp(\omega_k^T \Phi_k(\mathrm{x}_k)) \tag{10}$$

and divide function is

$$Z = \sum_k \exp\left(\sum_k \omega_k^T \Phi_k (\mathrm{x}_k) \right) \tag{11}$$

It can be found that the joint probability of the transition probability and the performance probability. The statistical probability is the conditional probability. Because it is only normalized locally, it is easy to fall into the local optimum.

4. Conclusion

In summary, the graph makes the probabilistic model visual, so that the relationship between some variables can be easily observed from the graph; at the same time, some complex computations can be understood as information transfer on the graph. This is where we don't have to pay attention to too many complex expressions. Finally, the graph model can be used to design new models. Of course, we can also consider its rationality from another angle. The model is the representation of the relationship between the two, and the probability theory is applied in the processing of the data. And the map just links the two together and plays a very good role.

References

[1] S. H. Strogatz: Nature, Vol. 410 (2001) No.6825, p.268.

[2] D. Napoletani and T. D. Sauer: Phys. Rev. E, Stat. Phys. Plasmas Fluids Relat. Interdiscip. Top., Vol. 77 (2008) No.2, p.1.

[3] B. Barzel, Y. Y. Liu and A. L. Barabási: Nat. Commun., Vol. 6 (2015) No.5, p.1.

[4] W. X. Wang, Y. C. Lai and C. Grebogi: Phys. Rep., Vol. 644 (2016) No.6, p.1.

[5] Y. Y. Liu, J. J. Slotine and A. L. Barabási: Nature, Vol. 473 (2011) No.7346, p.167.

[6] H. G. Tanner: *Proc. 43rd IEEE Conf. Decision and Control* (Bahamas December 14-17, 2004), Vol. 3, p.2467.

[7] N. Cai, J. X. Xi, and Y. S. Zhong et al.: Int. J. Innov. Comput. I., Vol. 8 (2012) No.5a, p.3315.

[8] N. Cai, J. W. Cao and M. J. Khan: Int. J. Control Automat. Syst., Vol. 12 (2014) No.6, p.1366.

[9] B. Liu, T. G. Chu, and L. Wang et al.: IEEE Trans. Autom. Control, Vol. 53 (2008) No.4, p.1009.

[10] Z. J. Ji, H. Lin, and H. S. Yu: Syst. Control Lett., Vol. 61 (2012) No.9, p.918.

[11] Z. J. Ji, H. Lin and H. S. Yu: IEEE Trans. Automat. Control, Vol. 60 (2015) No.3, p.781.

[12] J. H. Zhou, J. Z. Sun, K. Athukrala, D. Wijekoon and M.Ylianttila: Journal of Ambient Intelligence and Humanized Computing, Vol. 3 (2012) No.2, p.153.

[13] T. K. Shih and J. Vassileva: IEEE Transactions on Learning Technologies, Vol. 7 (2014) No.3, p.205.

[14] X. L. Liu, Y. J. Hsieh, R. Q. Chen, S. M. Yun: Complexity, DOI: 10.1155/2018/2170585.

[15] X. K. Liu, Y. Y. Ge, Y. Li: Complexity, DOI: 10.1155/2018/1487134.

[16] Q. Ye, Z. X. Jiang, T. N. Chen: Complexity, DOI: 10.1155/2018/5431987.

[17] Y. Z. Wang: Complexity, DOI: 10.1155/2018/7692375

[18] L. X. Gao, H. Fang, W. H. Chen, H. Cao: Complexity, Vol. 2019, DOI: 10.1155/2019/4583465.