

Bayesian Network Structure Learning from Attribute Uncertain Data

Wenting Song^{1,2}, Jeffrey Xu Yu³, Hong Cheng³, Hongyan Liu⁴,
Jun He^{1,2,*}, and Xiaoyong Du^{1,2}

¹ Key Labs of Data Engineering and Knowledge Engineering, Ministry of Education, China

² School of Information, Renmin University of China

{songwt, hejun, duyong}@ruc.edu.cn

³ The Chinese University of Hong Kong

{yu, hcheng}@se.cuhk.edu.hk

⁴ Department of Manage. Sci. & Eng., Tsinghua University

hyliu@tsinghua.edu.cn

Abstract. In recent years there has been a growing interest in Bayesian Network learning from uncertain data. While many researchers focus on Bayesian Network learning from data with tuple uncertainty, Bayesian Network structure learning from data with attribute uncertainty gets little attention. In this paper we make a clear definition of attribute uncertain data and Bayesian Network Learning problem from such data. We propose a structure learning method named *DTAU* based on information theory. The algorithm assumes that the structure of a Bayesian network is a tree. It avoids enumerating all possible worlds. The dependency tree is computed with polynomial time complexity. We conduct experiments to demonstrate the effectiveness and efficiency of our method. The experiments show the clustering results on uncertain dataset by our dependency tree are acceptable.

Keywords: uncertainty, Bayesian Network structure, dependency.

1 Introduction

In the past, researchers in data mining and machine learning area usually assume data is certain or precise, however that is not always the case. Data uncertainty arises in many applications such as sensor network and user privacy protection. Uncertain data can be divided into two categories by uncertainty source: artificial uncertain data and inherent uncertain data. People sometimes add noise to data for some purpose such as user privacy protection. As a result, the data become artificially uncertain. There are also inherently uncertain data. For example the scientific measurement techniques and tools are inherently imprecise and they are responsible for the generation of inherent uncertain data. Many researchers focus on uncertain data management and mining in database area and data mining area in recent years. In database area, there are three

* Corresponding author.

models of uncertain data. The first one is tuple uncertainty model [1] [2]. Each tuple in a probabilistic database is associated with a probability which represents the likelihood the tuple exists in the relation. The second one is attribute uncertainty model [3]. In attribute uncertainty model, each attribute in a tuple is subject to an independent probability distribution. Correlated uncertainty model [5] is the third one. Attributes are described by a joint probability distribution.

Many data mining algorithms have been proposed to analyze uncertain data, for example, mining frequent patterns from uncertain transaction database [2], naïve Bayesian classifiers for correlated uncertain data [7] [8], and clustering uncertain objects [6]. However there are few works on data mining from attribute uncertain data, and to the best of our knowledge no work has focused on how to learn Bayesian Network (*BN*) structure from such data. Attribute independency is a common assumption in database and data mining area, but it is not always reasonable, because there are dependency relationships among attributes. Attribute uncertainty due to measurement error or inherent uncertainty shouldn't be the reason for independence assumption. The structure learning from attribute uncertain data can reveal the essential relationship between attributes.

In this paper, we propose the problem of *BN* structure learning from attribute uncertain data and an algorithm named *DTAU* to solve the problem. Experiments demonstrate the effectiveness of our proposed algorithm.

The rest of the paper is organized as follows. Section 2 discusses related work on Bayesian Network structure learning for uncertain data. Section 3 gives relevant definitions. Section 4 introduces our structure learning algorithm *DTAU*. Section 5 is experimental study and Section 6 concludes the paper.

2 Related Work

Bayesian Network (*BN*) is a powerful tool to represent joint probability distribution over a set of variables or attributes. A *BN* is made up of two components: a directed acyclic graph (*DAG*), whose nodes represent variables and a set of conditional probability tables (*CPTs*) which specifies the conditional distribution of each variable given its parent in the *DAG*. Given a *BN* structure (*DAG*), there have been many algorithms for parameter learning. However, sometimes the *BN* structures are unknown for lack of domain knowledge. Thus the *BN* learning problem is of great importance. It has been proved that *BN* structure learning problem is *NP* completed [9].

Many researchers have proposed approximation algorithms to solve the structure learning from certain data problem. These methods are divided into two categories. The methods in the first category are based on information theory which is used to measure dependency relationships between nodes [10]. The methods [11] [12] in the second category aim to maximize score function of the possible structure considering that each node has no more than K parents. As this problem is *NP*-hard when K is bigger than 1, heuristic rules based methods are usually used. For the attribute uncertain data, we make use of the information theory to solve the problem and assume the structure is a tree.

3 Problem Definition

In this section we describe some concepts about the problem of learning *BN* structure from attribute uncertain data. The term *observation* is a concept in *BN* learning from certain data problem.

Definition 1 (Attribute). An *attribute* X_i is a component or aspect of an *object* O . X_i can take any value in $D(X_i)$ which is the possible value domain of X_i . D_i represents the size of $D(X_i)$. The *attribute* X_i is represented by a node (a random variable) in the *BN* structure.

Definition 2 (uncertain example). An vector $ue = \{P_1(X_1), P_2(X_2), \dots, P_m(X_m)\}$ is an *uncertain example* if each $P(X_i)$ is an probability distribution or probability density function over $D(X_i)$.

Definition 3 (Uncertain observation). Given an *attribute* X_i , an *observation* $P_i(X_i)$ of X_i ($1 \leq i \leq m$) in an uncertain example ue is an uncertain observation.

Definition 4 (Attribute Uncertain training dataset). An uncertain training dataset D is composed of uncertain examples, $D = \{ue_1, ue_2, \dots, ue_n\}$.

In this paper we focus on the problem of learning a *BN* structure from attribute uncertain training dataset. We assume the structure of Bayesian Network is a tree.

In the following parts of the paper we study the problem of learning structure from discrete attribute uncertain data. If they is continuous, we can discrete them.

4 Bayesian Network Structure Learning Algorithm DTAU

In this section, we start with a brief introduction of a naïve *BN* structure learning algorithm based on the exponential possible worlds. Then we will explain why the naïve method is unacceptable. At last we will show our approximation method which takes polynomial time.

Definition 5 (Possible world). Given an uncertain dataset about m attributes, it generates possible worlds, where each world is a certain dataset about the m attributes which has the same size of examples with the uncertain one. Each *possible world* W_i is associated with a probability $Pr(W_i)$ that the world exists.

The naïve method is based on possible world over all attributes (PWAA). The idea is converting the attribute uncertain dataset to some certain datasets. Given an uncertain discrete dataset AUD with N examples and m attributes, first we compute every possible world W_i of AUD . Each observations of example e_{ij} in possible world W_i is a possible value in the domain of the corresponding attribute. Those possible worlds form a set W and its size is $\prod_{i=1}^m (D_i)^N$. Second, we treat each possible world W_i as a certain training dataset, and then we learn a dependency tree under the corresponding training data set. The tree with the highest score can be recognized as the right one. The total number of trees (obviously containing the duplicates) is the number of the possible worlds. The score of a tree T_i is $\sum_{W_j.tree=T_i} Pr(W_j)$.

The number of possible worlds is exponential, so the solution presented above costs exponential time complexity to construct the dependency tree. We propose an algorithm named *DTAU* (Dependency Tree learning from Atttribute Uncertain data) to construct a dependency tree without enumeration of possible worlds and reduce the enormous computation. Our idea is to make use of the attribute uncertain dataset directly.

For a traditional certain training dataset, a popular way to construct the dependency tree with the closest probability approximation is called Chow-Liu tree [10]. The kernel idea in [10] is how to compute the dependency between each two attributes. The dependency between nodes X_i and X_j is measured by mutual entropy $I(X_i, X_j)$. The computation is defined by Equation 1. The value of mutual entropy $I(X_i, X_j)$ is always positive or zero. The value is more close to zero, then the dependency between the two attribute is weaker. The zero value means they are independent. We propose the *DTAU* algorithm to learn a dependency tree from attribute uncertain data. The *DTAU* algorithm is consistent with Chow-Liu tree under certain training data. The key point in the *DTAU* algorithm is how to compute the dependency between each two attributes under attribute uncertain training data. Equation 2 shows an initial approximation of $I(X_i, X_j)$. The two equations are from [10].

$$I(X_i, X_j) = \sum_{x_i, x_j} P(X_i = x_i, X_j = x_j) \log \frac{P(X_i = x_i, X_j = x_j)}{P(X_i = x_i)P(X_j = x_j)} \tag{1}$$

$$I(X_i, X_j) \approx \sum_{s \in D(X_i), t \in D(X_j)} N(X_i = s, X_j = t) \log \frac{N(X_i = s, X_j = t)}{\frac{n}{N(X_i = s)} \frac{n}{N(X_j = t)}} \tag{2}$$

Equation 3, 4 and 5 shows how to approximate the frequency in equation 2 and we get the final approximation of $I(X_i, X_j)$ by equation 3, 4 and 5.

$$N(X_i = s, X_j = t) = \sum_{k=1}^n P_{ki}(X_i = s) P_{kj}(X_j = t) \tag{3}$$

$$N(X_i = s) = \sum_{k=1}^n P_{ki}(X_i = s) \tag{4}$$

$$N(X_j = t) = \sum_{k=1}^n P_{kj}(X_j = t) \tag{5}$$

From the equations above we can learn that if the probability $P_{ki}(X_i=s)$ is the highest for attribute X_i and the probability $P_{kj}(X_j = t)$ is the highest for attribute X_j , then the occurrence probability for pair (s, t) may be the highest. The idea behind the equation is that the independence assumption doesn't have effect on the overall dependency computation. In other words, if the two attributes are independent, the computation result is zero. If they are not, the computation result is positive. The equation shows the consistence with certain data. We prove the result and we don't describe the details of the proof for the limitation of space in this paper.

The *DTAU* algorithm is divided into three steps. The first step is to compute the dependency between each two uncertain attributes and construct a weighted undirected graph. Then we follow the tree construction method in the Chow-Liu tree algorithm. The second step is to get a maximum spanning tree by a greedy algorithm

which is 2-ratio approximation of the optimal tree. The last step is to add the direction by width first traverse. The pseudo-code is given below:

Algorithm DTAU

Input: an attribute uncertain dataset AUD with n examples and m variables

Output: a dependency tree T

Major steps:

Construct a weighted undirected completed graph $G = (V, E)$, $V = \{X_1, X_2, \dots, X_m\}$;

for $i = 1$ to m

 for $j = 1$ to i

 Compute dependency $I(X_i, X_j)$ by equation 1, 2, 3, 4 and 5;

$e(X_i, X_j) = I(X_i, X_j)$

 end for

end for

$T = \text{greedy_max_spanning_tree}(G)$;

Select a random node X_x in T as beginning node;

Add direction for each edge by breadth-first traverse beginning with X_x ;

The time complexity of DTAU algorithm is $O(nm^2)$, which is smaller than the naïve solution. We can confirm that if the difference between $I(X_i, X_j)$ and $I(X_i, X_k)$ satisfies a t -condition that the absolute value of the difference between $I(X_i, X_j)$ and $I(X_i, X_k)$ is bigger than t , the dependency tree created by the DTAU method is the same with the one in the possible world with the highest probability. Because if the t -condition is satisfied, the partial orders for all mutual entropy are the same. The partial orders can determine the structure. We prove this result and we don't describe the details of the proof or the computation of t for the limitation of space. Experiments show that our method performs well even when the t -condition can't be satisfied.

5 Experiments

As there hasn't been any public attribute uncertain dataset, the attribute uncertain datasets we use are generated from certain datasets artificially. We generate the attribute uncertain datasets by adding noise to the UCI machine learning datasets which are standard for traditional BN learning problems. We convert the Letter recognition and Balance datasets to attribute uncertain datasets. The noise addition strategy is described as follows. For each training example in the original certain dataset, we assign a probability p which is not smaller than a bound α to the corresponding attribute's observation in the original certain training example, and assign a low probability to other possible values for this attribute. Table 1 shows an example of the original certain dataset, where $D(\text{Attribute } 1) = \{a, b, c\}$ and $D(\text{Attribute } 2) = \{d, e, f\}$. Table 2 shows an uncertain dataset obtained after noise addition with α being 0.5. By this way we get attribute uncertain training datasets denoted by AU-Letter- α , and AU-Balance- α respectively.

Table 1. An example of certain data

Attribute 1	Attribute 2
<i>a</i>	<i>d</i>
<i>b</i>	<i>e</i>

Table 2. An example of noise addition

Attribute 1	Attribute 2
<i>a</i> : 0.8 <i>b</i> : 0.1 <i>c</i> : 0.1	<i>d</i> : 0.5 <i>e</i> : 0.25 <i>f</i> : 0.25
<i>a</i> : 0.2 <i>b</i> : 0.6 <i>c</i> : 0.2	<i>d</i> : 0.15 <i>e</i> : 0.7 <i>f</i> : 0.15

For each uncertain dataset we generate, the uncertain observations in an attribute uncertain dataset are closer to the ones of the original certain dataset when α is closer to 1. The experiment on AU-Letter-0.5 shows that the partial orders of the dependency measure (information entropy) in attribute uncertain data is almost the same with the one in the certain dataset. Figure 1 shows the tree we learn from AU-Letter-0.5 and the Chow-Liu tree learned from the certain dataset.

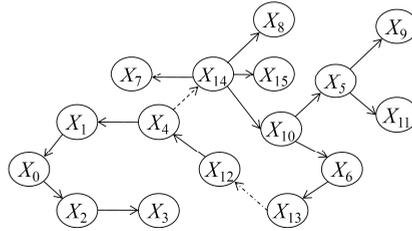


Fig. 1. Dependency tree from the dataset Letter and the correspondingly certain dataset

The two trees share the black solid edges. Edge $\langle X_{13}, X_{12} \rangle$ belongs to the tree from uncertain dataset and edge $\langle X_4, X_{14} \rangle$ belongs to the tree from certain dataset. We find that only edge $\langle X_4, X_{14} \rangle$ and edge $\langle X_{13}, X_{12} \rangle$ are different. The difference between $I(X_4, X_{14})$ and $I(X_{12}, X_{13})$ accounts for the bigger one of the two information entropy less than 1.3%.

We design experiments on AU-Balance to demonstrate the effectiveness of the dependency tree by clustering results. We use *DTAU* algorithm to learn a dependency tree from AU-balance data. Then we generate a certain sample dataset i for the uncertain training example ue_i in the dataset. Then we treat the sample dataset i as the training dataset and the dependency tree as *BN* structure to learn the joint probability distribution on all attributes in the uncertain dataset and then the uncertain example ue_i turns to be an uncertain object. We cluster those uncertain objects by algorithm *UK-means* [4] and the original certain objects by *K-means* and compare the two clustering results. We do experiments on the AU-Balance dataset with different values of α .

Table 3. Clustering precision under different parameters

Dataset	Cluster 0	Cluster 1	Cluster 2	precision
AU-Balance-0.6	6	161	103	43.2%
AU-Balance-0.8	21	115	98	37.44%
AU-Balance-0.9	5	119	156	44.8%
Balance certain	16	175	145	53.74%

We test three different values of parameter α , 0.6, 0.8 and 0.9. For each dataset, the cluster result is compared with the true class labels. Table 3 shows the results under different uncertain dataset. The numbers in column 2, column 3 and column 4 represent the size of correct examples in the corresponding cluster. The measure precision shows the percentage of correct clustered examples. From this table we can see that the precision for each of the three uncertain dataset is quite close to the certain one, for the certain one is always the possible world with the highest probability. The experiments show that the dependency tree generated by our method is acceptable and α is an important factor to the cluster results.

6 Conclusion

In this paper we propose the Bayesian Network structure learning problem on attribute uncertain dataset, and we propose algorithm *DTAU* by which we can learn a dependency tree in polynomial time. We conducted experiments to demonstrate the effectiveness of our proposed algorithm. The experiment results show the dependency trees are acceptable and the proposed algorithm is effective. In the future, we will further analyze the effect of parameters α .

Acknowledgement. This work was supported by the NSFC under Grant No. 70871068 and 71110107027 and the Major National Sci. and Tech. Project of China under Grant No. 2010ZX01042-002-002-03.

References

1. Dalvi, N., Suciu, D.: Efficient query evaluation on probabilistic databases. *The VLDB Journal* 16(4), 523–544 (2007)
2. Bernecker, T., Kriegel, H.P., Renz, M., Verhein, F., Zuefle, A.: Probabilistic frequent item set mining in uncertain databases. In: 15th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, pp. 119–128. ACM Press, Paris (2009)
3. Singh, S., Mayfield, C., Shah, R., Prabhakar, S., Hambrusch, S., Neville, J., Cheng, R.: Database support for probabilistic attributes and tuples. In: 24th IEEE ICDE International Conference on Data Engineering, pp. 1053–1061. IEEE Press, Cancún (2008)
4. Lee, S.D., Kao, B., Cheng, R.: Reducing UK-means to K-means. In: 7th IEEE ICDM Workshops International Conference on Data Mining Workshops, pp. 483–488. IEEE Press, Omaha (2008)
5. Gullo, F., Ponti, G., Tagarelli, A., Greco, S.: A Hierarchical Algorithm for Clustering Uncertain Data via an Information-Theoretic Approach. In: 8th IEEE ICDM International Conference on Data Mining, pp. 1053–1061. IEEE Press, Pisa (2008)
6. Günemann, S., Kremer, H., Seidl, T.: Subspace Clustering for Uncertain Data. In: SIAM SDM SIAM Conference on Data Mining, pp. 385–396. SIAM Press, Ohio (2010)
7. He, J., Zhang, Y., Li, X., Wang, Y.: Naive Bayes classifier for positive unlabeled learning with uncertainty. In: SIAM SDM SIAM Conference on Data Mining, pp. 361–372. SIAM Press, Ohio (2010)

8. Ren, J., Lee, S.D., Chen, X., Kao, B., Cheng, R., Cheung, D.: Naive bayes classification of uncertain data. In: 9th IEEE ICDM International Conference on Data Mining, pp. 944–949. IEEE Press, Miami (2009)
9. Dalvi, N., Suciu, D.: Learning Bayesian networks is *NP*-complete. Lecture Notes In Statistics, pp. 121–130. Springer, New York (1996)
10. Chow, C., Liu, C.: Approximating discrete probability distributions with dependence trees. IEEE Transactions on Information Theory Journal 14(3), 462–467 (1968)
11. Heckerman, D., et al.: A tutorial on learning with Bayesian networks. In: Learning in Graphical Models, Michael I. Jordan, Massachusetts (1999)
12. Friedman, N., Nachman, I., Peér, D.: Learning Bayesian Network Structure from Massive Datasets: The “Sparse Candidate” Algorithm. In: UAI 14th Conference on Uncertainty in Artificial Intelligence, Madison, pp. 206–215 (1999)