SOIL FERTILITY GRADING WITH BAYESIAN NETWORK TRANSFER LEARNING

HAI-YANG JIA[1], JUAN CHEN[1], HE-LONG YU[1,2], DA-YOU LIU[1]

[1] College of Computer Science and Technology ; Key Laboratory for Symbolic Computation and Knowledge Engineering of Ministry of Education, Jilin University, Changchun , Jilin China 130012
[2] College of Information technology, Jilin agricultural University, Changchun, Jilin China 130118
E-MAIL: chenjuan.jlu.edu.cn

Abstract:

Soil fertility grading is an important issue in the agriculture domain, AI based approach has been applied recently. But in most circumstance data obtaining is a expensive and time consuming procedure, sometimes even impossible. This paper presents a Bayesian Network based transfer learning algorithm. The existing training results can be transferred between the nearby land squares. The proposed algorithm considers both the similarity between the learning task and the geographical position of the land squares. Empirical experiment were implemented to prove the efficiency of the algorithm.

Keywords:

Soil fertility grading; Bayesian network; Transfer learning; Machine learning

1. Introduction

Soil fertility grading is a important issue in agriculture field, many artificial intelligent based approach, such as Bayesian network, artificial neural network and D-S theory [1], has been applied recently. Most of the existing algorithm assumed that a sufficiently large dataset is available from which a reliable model can be induced. But in the context of soil fertility grading, train and test data obtaining about the soil is very expensive and time consuming procedure, sometimes even impossible.

Recently, there has been an increasing interest in the machine learning community for using data from related tasks, in particular when the available data is scarce [2-4]. In such cases, knowledge transfer or transfer learning between task domains would be desirable. This paper proposes a Bayesian Network based transfer learning algorithm. The exits training result can be transferred between the nearby land squares. The transfer algorithm considers both the similarity between the learning task and the geographical position of the land square.

The remainder of this paper is organized as following: Section 2 introduces background knowledge; Section 3 describes the algorithm in detail; section 4 describes the empirical experiment and finally section 5 draw the conclusion.

2. Background

Notation: capital letter X,Y,Z notate a random variable; lowercase x,y,z notate the value of a random variable; bold capital letter **X,Y,Z** notate a set of random variable.

2.1. Bayesian Network Learning

A Bayesian network is a concise representation of joint probability distribution on a set of random variables [5].

Definition1: A *Bayesian Networks(BNs)* consists of two parts $BNs = < G, \theta >$ (qualitative and quantitative part), represent the joint probability distribution of a set of random variables **X**=(X_1,X_2,…X_n):

G is a directed acyclic graph, short for DAG, each node corresponds to a random variable in X, and G encodes the (condition) independencies of the probability distribution.

θ is conditional probability table, short for CPT, encoding the conditional distributions of each family (a node and its parent node), $\theta = \{p(X_i \mid \pi_{X_i}) \mid 1 \le i \le n\}$ (π_{X_i} is the parent nodes of X_i).

Briefly, the joint probability distribution represented by BNs is:

$$P(X_1, X_2, ..., X_n) = \prod_{i=1}^{n} P(X_i \mid \pi_{X_i})$$

Fig 1 gives a BNs with 6 random variables: B E A R J M, each variable has two state (T F). Joint probability distribution is (each conditional distribution can be found in CPT):

$$P(B,E,A,R,J,M) =$$
$$P(B)P(E)P(A \mid E,B)P(R \mid E)P(J \mid A)P(M \mid A)$$

978-1-4244-6527-9/10/$26.00 ©2010 IEEE

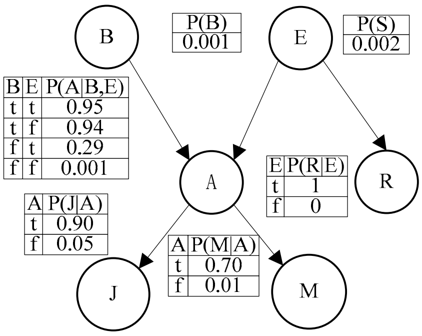

Figure 1. BNs about earthquake

Learning a BN includes two aspects: learning the structure and learning the parameters. When the structure is known, parameter learning consists on estimating the CPTs from data. Structural learning is the key component of BNs learning. There are two approaches for BNs structural learning: One is score based approach, which define the task as an optimization problem. Using a scoring function score (G, D) to evaluates different networks G relative to a data set D (in the rest of this paper, omit D for mentioning it explicitly). There are many score criteria such as: BD/BDe [6, 7], minimum description length, MDL [8] and Bayesian Information Criterion, BIC [9]; the other approach is constraint based approach, which defines this task as constraint satisfaction problem. Applying conditional independent test to find the conditional independence relationships in data D, then construct a BN satisfied such constraint (conditional independence) [10]. Each approach has its advantage and disadvantage: the constraint based methods are usually much more efficient when the number of variables is large. However, when the sample size is small and the data is noisy, the scoring based algorithms can often more accurate. Both structure learning approach, including their variants; require enough data to produce accurate results.

2.2. Transfer Learning

Traditional data mining and machine learning algorithms make predictions on the future data using statistical models that are trained on previously collected labeled or unlabeled training data. Most of the existing approaches assume that the distributions of the labeled and unlabeled data are the same. Transfer learning, in contrast, allows the domains, tasks, and distributions used in training and testing to be different. The study of Transfer learning is motivated by the fact that people can intelligently apply knowledge learned previously to solve new problems faster or with better solutions. Research on transfer learning has attracted more and more attention, such research are presented (or closely related to) with different names: learning to learn, life-long learning, knowledge transfer, inductive transfer, multi-task learning, knowledge consolidation, context sensitive learning, knowledge-based inductive bias, meta learning, and incremental/cumulative learning [11, 12].

2.3. Relate Approaches

Recently there has been an increasing interest in transfer learning for different types of representations. Dai et al [13] proposed a transfer learning algorithm for text classification. Roy and Kaelbling [14] developed an alternative method for transfer learning for the naïve Bayesian classifier. Richardson and Domingos [15] consider the combination of multiple experts, and develop a Bayesian approach in which the knowledge of several experts is encoded as the prior distribution of possible structures, and the data is used to refine this structure and learn the parameters. Niculescu and Caruana [16] consider the problem of learning Bayesian network structures for related tasks. Luis et al [17] based on independence tests that are obtained separately for each dataset, and then combined, resulting in a simpler method, learn simultaneously the models for different problems.

Previous approaches for transfer learning BNs consider the combination of data or expert knowledge to learn simultaneously several related tasks solves a different although related problem. In contrast, our approach focuses on the soil fertility grading issue, in this scenario geographical position of the land square is another important feature. The algorithm considers both the similarity between learning tasks and the geographical position of the land square. The algorithm as described below

3. Learning Algorithm

BNs learning include structure learning and parameter learning. There are two approaches for Traditional BNs structural learning:

- One is Score based approach, which define the task as an optimization problem. Define a scoring function score (G, D), which evaluates different networks relative to the data D (in the rest of this paper, omit D mentioning it explicitly). There are many score criteria such as: BD/BDe [6, 7], minimum description length, MDL [8] and Bayesian Information Criterion, BIC [9];
- The other is constraint based approach, which define this task as constraint satisfaction problem. Applying conditional independent test to find the conditional independence relationships in data D, then find BNs satisfied such constraint (conditional independence)

[10].

Each approach has its advantage and disadvantage; the constraint based methods are usually much more efficient when the number of variables is large. However, when the sample size is small and the data is noisy, the scoring based algorithms can often give more accurate. In this paper, we combine these approaches.

3.1. Structural learning

Let us assume that the soil fertility grading task for number k land square represented by a set of discrete random variables: $\mathbf{X}^k=(X_1,X_2,\ldots X_n)$, there are n land squares, represented by $\mathbf{X}^k$ (k=1..n), training data sets are D^k (k=1..n). The target land square, number t square($1\leqslant t\leqslant n$), have a very small training data set D^t, except the target square, the rest square, called auxiliary square, have sufficient data for learning BNs. The objective of the learning algorithm is to build a BNs for the target task, BN^t, from dataset D^k (k=1..n).

Phase 1: Mutual information and the conditional mutual information are employed to perform conditional independence test (CI test), for $k\in[1,n], k\neq t$:

$$I_k(A,B)=\sum_{a,b}P_k(a,b)\log\frac{P_k(a,b)}{P_k(a)\,P_k(b)}$$

$$I_k(A,B\mid C)=\sum_{a,b,c}P_k(a,b,c)\log\frac{P_k(a,b\mid c)}{P_k(a\mid c)\,P_k(b\mid c)}$$

It is supposed that A is (conditional) independent of B when $I<\varepsilon$, short for Ind (A, B) and Ind (A, $B|C$). for k=t

$$I_t(A,B)\ =\sum_{k=1}^{n,k\neq t}\alpha_k(A,B)I_k(A,B)$$

$$I_t(A,B\mid C)\ =\sum_{k=1}^{n,k\neq t}\alpha_k(A,B\mid C)I_k(A,B\mid C)$$

Where α_k is the weight factor of land square, which depends on the following aspects: 1) the size of number k dataset 2) similarity of distribution between target and auxiliary dataset 3) the geographical position of the land square. In the following equation, M,S and W represent the aspects above:

$$\alpha_k(A,B\mid C)=M_k(A,B\mid C)\times S_k(A,B\mid C)\times W_k$$

The influence from the size of the dataset can empirically be shown that the error of this test is asymptotically proportional to the following equation[18] (where L is the size of the dataset, |A| is the size of A):

$$M_k(A,B\mid C)=(1-\frac{\log L}{2L})\times|A|\times|B|\times|C|$$

Similarity of distribution between target and auxiliary dataset defined by the product of global similarity S_k^g and local similarity $S_k^l(A,B\mid C)$ (we follow the definition in paper [17]):

$$S_k(A,B\mid C)=S_k^g\times S_k^l(A,B\mid C)$$

The geographical position of the land squares are presented with polar coordinate system, target land squares $p_t=(\rho_t,\theta_t)$, auxiliary land squares $p_k=(\rho_k,\theta_k)$. The weight factor W related to the distance and relative orientation of the two squares (where $t(\theta)$ is a experiential function that quantize the relative orientation).

$$W_k=t(\theta_k-\theta_t)(\rho_t^2+\rho_k^2-2\rho_k\rho_k\cos(\theta_k-\theta_t))$$

Phase 2: in phase 1, we perform order 0 and order 1 CI test, then phase 2 is searching the best essential graph in a constrained searching space. Detail proof is in our previous work [19]. Here we the extend the score function to transfer learning scenario.

$$\mathrm{Sc}_t\left(\mathrm{G}\right)=\sum_{k=1}^{n,k\neq t}\alpha_{\mathrm{k}}\mathrm{Sc}_k\left(\mathrm{G}\mid\mathrm{D}^k\right)$$

$$Sc(DAG)=\sum_i(\sum_{l=1}^{M}\log_2P(X_i\mid\pi_{X_i})\text{-}\frac{1}{2}\|\pi_{X_i}\|(\|X_i\|-1)\log_2L)$$

where $\|X\|$ is the number of state of X, X_i^k is the k_{th} state of X_i.

3.2. Parameter learning

After learning a Bayesian network structure G_t, the next step is to complete the parameter, conditional probability tables (CPTs), for each variable given its parents need to be estimated. For this task, we apply instance transfer [2] to complete the parameter learning. For learning the parameter with target structure G_t, the data set from m most closest auxiliary land square (the land square that correspond to m highest α_k) were transferred. With these data traditional parameter learning algorithm can be applied, such as maximum likelihood estimation (MLE):

$$\theta^*=\arg\sup_\theta L(\theta\mid D)$$

$$\theta_{i,j,k}^*=\begin{cases}\dfrac{N_{i,j,k}}{N_{i,j}}, & \text{if} N_{i,j}>0\\ \dfrac{1}{r_i}, & \text{else}\end{cases}\text{, where }N_{i,j}=\sum_{k=1}^{r_i}N_{i,j,k}$$

$N_{i,j,k}$ represent the number of samples that satisfy $X_i=k$, $\pi_{X_i}=j$.

4. Experiments

The purpose of the experiments is to show how BNs learn from scarce data of the target land square can be improved with data from similar square. The data set is about the soil in NongAn China, number 1-7 land squares are concerned, following table is a slice of training data.

TABLE 1. A SLICE OF TRAINING DATA

Ph	Cu	Fe	K	Mn	N	P	Org	Zn	Gd
6.6	1.25	8.72	146.0	22.33	0.095	15.4	2.12	1.35	5
6.7	1.26	8.47	147.9	22.41	0.096	15.1	2.13	1.34	6
6.7	1.39	4.70	161.3	22.05	0.102	13.0	2.18	1.35	5
6.5	1.21	9.45	141.8	22.44	0.092	16.4	2.07	1.37	5
6.7	1.27	9.71	149.2	21.88	0.098	14.2	2.15	1.33	5
6.6	1.20	8.30	138.7	21.25	0.088	17.1	2.02	1.44	5
6.6	1.30	8.50	144.8	22.30	0.093	15.9	2.11	1.39	5
6.7	1.28	8.48	148.8	21.89	0.097	14.3	2.16	1.34	6
6.7	1.30	0.10	152.9	21.90	0.099	13.8	2.17	1.35	6

To compare the learning result, number 3 land square were selected as target one (assumed that the training data were scarce), transfer learning approach and traditional approach a performed, the result of the transfer algorithm showed in fig 2, the traditional approach find the structure that has 4 edges lost. According to the domain expert, the lost edges are necessary and the results are more accuracy.

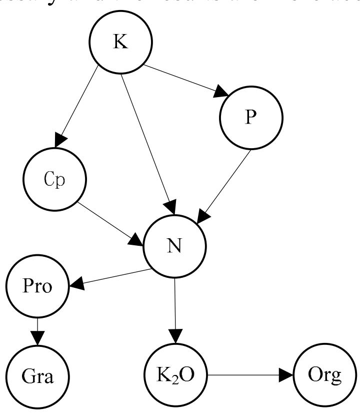

Figure 2. Learned structure

5. Conclusion

In this paper we have presented a novel methodology for transfer learning Bayesian networks for soil fertility grading problem. The algorithm including structure learning and parameter learning, considering both the similarity between learning task and the geographical position of the land square. Through using the information from related datasets, the performance of networks constructed with small datasets was improved. Empirical experiment results show a significant improvement in terms of structure and parameters when transfer knowledge between similar soil fertility grading task.

Future work, we would like to explore how to deal with incomplete data set and the situation that the auxiliary data are also scarce.

Acknowledgements

This paper is supported by National Natural Science Foundation of China under Grant Nos. 60773099, 60873149, 60973088, the National High-Tech Research and Development Plan of China under Grant 2006AA10Z245, 2006AA10A309, Research Fund for the Doctoral Program of Higher Education of China under Grant No 20070783057, the Special funds of Central Colleges Basic Scientific Research Operating Expenses, Jilin University under Grant No. 421032041421.

Corresponding author: Juan Chen, chenjuan@jlu.edu.cn, Jilin University, 2699 Ave. Qianjin, Changchun, 130012, P.R.China.

References

[1] 2004, "The Research and Implementation of the Inference Tools Based on D-S Theory and Convex Evidence Theory," Master, Jilin university, shudan zhang.

[2] S. J. Pan and Q. Yang, "A Survey on Transfer Learning," *IEEE Transactions on Knowledge and Data Engineering,* vol. 99, 2009.

[3] D. Silver and K. Bennett, "Guest editor's introduction: special issue on inductive transfer learning," *Machine Learning,* vol. 73, pp. 215-220, 2008.

[4] R. Caruana, "Multitask Learning," *Machine Learning,* vol. 28, pp. 41-75, 1997.

[5] J. Pearl, *Probabilistic reasoning in intelligent systems: networks of plausible inference*. San Mateo, California: Morgan Kaufmann, 1988.

[6] H. Cooper, "A Bayesian Method for the Induction of Probabilistic Networks from Data," *Machine Learning,* vol. 9, pp. 309-347, 1992.

[7] H. David*, et al.*, "Learning Bayesian networks: The combination of knowledge and statistical data," *Machine Learning,* vol. V20, pp. 197-243, 1995.

[8] B. F. Lam W " Learning Bayesian Belief Networks: An Approach Based on the MDL Principle," *Computational Intelligence,* vol. 10, pp. 269-293, 1994.

[9] Schwarz, "Estimating the dimension of a model,"

Annals of Statistics, vol. 6, pp. 461-464, 1978.
[10] J. Cheng*, et al.*, "Learning Bayesian networks from data: An information-theory based approach," *Artificial Intelligence,* vol. 137, pp. 43-90, May 2002.
[11] M. M. H. Mahmud, "On universal transfer learning," *Theoretical Computer Science,* vol. 410, pp. 1826-1846, Apr 28 2009.
[12] S. Thrun and L. Pratt, *Learning to learn*. Norwell, MA, USA:: Kluwer Academic, 1998.
[13] W. Dai*, et al.*, "Trasfer naive Bayes classifiers for text classification," presented at the AAAI-07, 2007.
[14] D. M. Roy and L. P. Kaelbling, "Efficient Bayesian task-level transfer learning," presented at the Proceedings of the 20th international joint conference on Artifical intelligence, Hyderabad, India, 2007.
[15] M. R. Mattr and P. Domingos, "Learning with Knowledge from Multiple Experts," presented at the ICML 2003, 2003.
[16] N. M. Alexandru and R. Caruana, "Inductive Transfer for Bayesian Network Structure Learning," presented at the Eleventh International Conference on Artificial Intelligence and Statistics, 2007.
[17] R. Luis*, et al.*, "Inductive transfer for learning Bayesian networks," *Machine Learning,* vol. 79, 2010.
[18] N. Friedman and Z. Yakhini, "On the sample complexity of learning Bayesian networks," presented at the Proceedings of the 12th conference on Uncertainty in Artificial Intelligence, San Mateo, 1996.
[19] J. Haiyang*, et al.*, "A Hybrid Approach for Learning Markov Equivalence Classes of Bayesian Network," presented at the International Conference on Knowledge Science, Engineering and Management 2007, Melbourne, Australia, 2007.