A Method for Predicting Understanding of Course Knowledge based on Junction Tree

Lijin Long, Huyi Liang

Department of Electronic and Information Engineering, Zhejiang University of Media and Communications, Hangzhou, 310018, China

email: zj_education_ljl@163.com

Keywords: course inference model; knowledge points; prior knowledge; junction tree

Abstract. For the same course, different application fields have different requirements for course content, and their course syllabuses cover knowledge points all. Select some knowledge points from course syllabus to create interrelationship among them based on knowledge structure, and generate conditional probability tables according to prior knowledge of course teaching experts so that the Bayesian network of course knowledge points can be created to show field knowledge coverage and knowledge inference among these nodes existing in the Bayesian network. In order to update nodes by message passing conveniently, a Bayesian network need to be transformed into a junction tree, which can be trained by historical course data recorded by course teachers, to improve accuracy of inference for subsequent knowledge learning effect based on learned knowledge points.

Introduction

Course syllabus covers knowledge points all and corresponding demands for course teaching. There exists the logical relationship among these knowledge points so it can be transformed into a logical diagram. In the logical diagram, nodes, also called as random variables, their weight values can be quantified by prior knowledge of course teaching experts, so that logical inference model can be created based on logical diagrams and corresponding condition tables, which are generated by course teaching experts. The created logical inference model can infer study effect of subsequent course knowledge points according to learned course knowledge points for different students. If course inference model is created only by prior knowledge of course teaching experts to cover all course knowledge points and for all students, it will be a significant amount of work and it can't adapt to dynamic data either. Therefore, for this kind of application, there need to create a dynamic course inference model, that is, random variables can be inferred by message passing based on the trained course inference model and observations.

As one of uncertain inference models, Bayesian network has been applied into most of fields such as education, medical diagnosis, information retrieval, and so on. There are a lot of inference algorithms have been proposed to apply into Bayesian network, such as junction tree, clique tree propagation, figure reduction, poly-tree propagation, etc. For the application of course knowledge inference, the algorithm of junction tree has some advantages like easy analysis and understanding, high accuracy, efficient[1-4]. Therefore, this paper introduces how to create a Bayesian network based on course knowledge and prior knowledge of course teaching experts and how to transform the created Bayesian network into a junction tree so as to improve dynamic adaptability to update course inference model based on new data. As an application case of this course inference model, the knowledge points of network layer belonging to course "Data Communication and Network" are selected to create course inference model, and it is applied to predict the subsequent study effect on understanding of knowledge points by teachers or students based on prior knowledge of course teaching experts and historical course data, so as to meet different demands for course teaching.

Overview

The Bayesian network of local knowledge structure is commonly created by basic units such as course knowledge points, course teaching distributions, chapters and sections, and so on. Once node feature of Bayesian network has been determined, prior knowledge of course teaching experts can be applied into generating prior probability table and condition probability table, also named as PPT and CPT in brief, so as to reflect learners' mastery of course knowledge content. In a Bayesian network, every node has a level mapped to mastery of course knowledge, marked with A,B,C,D,E corresponding to excellent, good, medium, fair, poor. In order to update course inference model conveniently, there need to transform a Bayesian network into a junction tree, that is, the directed acyclic graphs, also called as DAG for short, should convert into the undirected graphs. After that, they are divided into some maximum connected sub-graphs, also called as cliques or classes, based on numbers and weight values of members in cliques. In a Bayesian network, DAG is viewed as a single class, then a maximum connected sub-graph is viewed as a clique or a subclass. According to numbers and weight values of members in a clique, the separated set can be generated to connect at least two nodes in a clique. When the message passing between two random nodes is unique, i.e., the separated set connecting these two nodes is unique, the course inference model reaches steady state and the junction tree reaches consistent state, which called the local optimized junction tree[5,6]. Before that, course historical data are used as observations to eliminate inconsistent junction tree by iterative optimal algorithm. According to different application fields, select different course knowledge points, which cover corresponding field application, and construct n local optimized junction trees based on the method stated before. These n junction trees are viewed as n cliques or clusters. By means of prior knowledge of course teaching experts, create the interrelationship and CPT among these cliques, and let every clique has an appropriate weight value, reuse the proposed method to construct a synthetic junction tree, which can be used to predict the subsequent study effect of course knowledge points based on learned course knowledge points so as to provide a practical foundation for students and teachers to improve or enhance study effect in the following course learning.

Constructing of Course Inference Model

A Bayesian network is composed of a set of conditional probability functions and a directed acyclic graph of uncertain inference model. In the application of course inference model, the Bayesian network is created by interrelation among selected course knowledge points for specified field application, and by prior knowledge of course teaching experts. As stated before, the proposed method suggests to transform a Bayesian network into a corresponding junction tree. How to implement to transform and how to get a local optimized junction tree are described as follows.
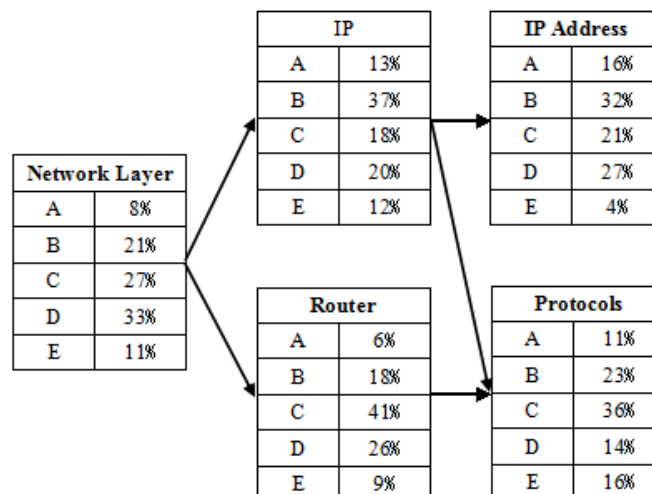
Network Layer	
A	8%
B	21%
C	27%
D	33%
E	11%

IP	
A	13%
B	37%
C	18%
D	20%
E	12%

IP Address	
A	16%
B	32%
C	21%
D	27%
E	4%

Router	
A	6%
B	18%
C	41%
D	26%
E	9%

Protocols	
A	11%
B	23%
C	36%
D	14%
E	16%

Figure.1 Prior knowledge of course teaching experts

Bayesian Network

The definition of a Bayesian network includes a DAG and a set of CPT. Every node in the DAG represents a random variable, which can be observed directly or be hide variable. Every directed edge in the DAG represents condition dependence between or among of random variables in the DAG. In condition probability table, every element is corresponding to a unique node, and every cell stores joint condition probability of all direct forwarding nodes. There exists an important property that the condition of every node is independent of all undirected forwarding nodes before its directed forwarding node is determined, which is similar with Markov process. It also can be viewed as non-linear extension of Markov chain. Based on the property, The joint probability distribution of Bayesian network can be calculated based on DAG and CPT. Eq.1 shows how to calculate joint condition probability distribution of Multi-variables and non-independent.

$$P(x_i) = \prod_{i=1}^{n} P(x_i \mid parents(x_i)) \square \quad \square\square\square$$

In (1), symbol "Parents" represents joint distribution of all directed forwarding nodes of x_i, and their probability values are calculated by prior condition probability of course teaching experts. The example of network layer is shown in Fig.1. If one node has not been pointed, it is called root node, and if has, it is called child node. The edges existing in Fig.1 represent one node has the influence on another node.

Junction Tree of Course Knowledge Points

The logical relationship between or among course knowledge points can be shown by a Bayesian network. However, the course inference model need to be constantly updated, when new valid data have been acquired. For this reason, the proposed method puts to use the inference model of junction tree, and eliminates inconsistent junction tree by iterative optimal steps to generate local optimized junction tree based on numbers and weight values of members in a clique. In a junction tree, every node is called as a clique and is corresponding to one or more than one random variables in Bayesian network. When it is corresponding to more than one random variables, it is composed of maximum connected sub-graphs in a Bayesian network. The node connected two or more than two other nodes is the intersection node, it is the key node to pass message in a Bayesian network. Assume C_i and C_j represent two different cliques, the node connected the two cliques is defined as : $I_{ij} = C_i \cap C_j$. The parameter of every clique is a distribution function, symbolized as F_C, which is calculated based on CPT. According to joint probability distribution defined in (2), F_C is calculated by selecting the result of greater or equal to zero from combination of random variables in a Bayesian network. The joint probability distribution is defined as:

$$P = \frac{\prod F_C}{\prod F_I} \square \qquad \square 2\square$$

In (2), F_C represents the probability distribution of intersection set between C_i and C_j. There are two kinds of F_C, the one is candidate separated set, the other is separated set. As stated in II, when a consistent junction tree meets the criteria $F_{C_i} = F_I = F_{C_j}$, the course inference model reaches the steady state. In the process of inference training, message passing between C_i and C_j is by means of I_{ij}. Every probability distribution of random variable in a Bayesian network is calculated by $P(V) = \sum_{C \setminus \{V\}} F_C$ In the same way, $\square P(V/E) = \frac{P(V = v \mid E = e)}{P(E = e)}$.

Some course knowledge points belonging to network layer in course "Data Communication and Network" is shown in Fig.2.
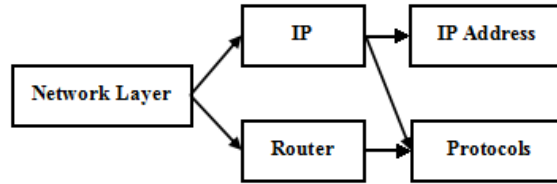
Figure.2 The logical relationship of course knowledge points

How to transform a logical relationship diagram into a local optimized junction tree, the procedure is described as follows. First, according to Fig.2, construct undirected graph G_u. Eliminate direction in Fig.2 and add an edge between parent nodes. Second, there are three specific steps.

a. Select a node x from G_u. In step b, the selection of node x makes minimum numbers of edge.

b. A non-empty set, the part of a clique, is composed of node x and its neighbor nodes. All nodes of connecting node x and its neighbor nodes are added in the set and edges connected two or more than two nodes are also added in a clique.

c. Remove node x from G_u. Repeat step a to step c until there is no change in G_u.

After above two steps, G_u is transformed into G_T shown in Fig.3. In G_T, every weight value of node is same as G_u and every weight value of a set is the product of all members.
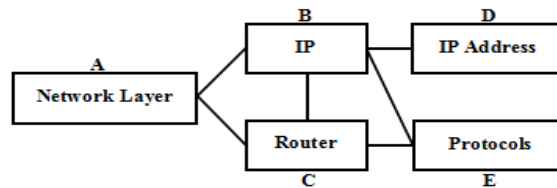

Figure.3 The transformation result of logical relationship

Third, according to Fig.3, the maximum connected sub-graphs, also called as cliques, can be separated from G_T, i.e., ABC, BCE, BD. Fourth, construct a local optimized junction tree. In order to improve inference performance after constructing a junction tree, there need to connect these cliques by separated sets so as to meet the optimality criterion. Separated sets come from candidate separated sets, and it defined as follows: in clique sets, if there exists the fact that member A reaches member C passes member B only, and vice versa, then member B is defined as a separated set. Member B is a clique, which has at least one or more than one nodes. In clique sets with n members, repeat the steps to add edges or separated sets until there are $n-1$ edges or separated cliques to connect all cliques. The specific steps of the algorithm are described as follows.

a. Initialize n cliques and a empty set I.

b. Select two cliques C_i and C_j randomly to generate a candidate separated set, denoted as $I_{ij} = C_i \cap C_j$.

c. Put I_{ij} in I and repeat above steps until there exists that $n-1$ separated sets have been insert into all clique sets.

Select a candidate separated set I_{ij} from I and search all paths from C_i to C_j to check whether they all pass I_{ij}, i.e., whether I_{ij} is a key clique to connect C_i and C_j. If so, the candidate separated set I_{ij} is selected to separated set directly. If not, select the candidate separated set with the maximum numbers of member, denoted as $N(I_{ij})$, and the minimum weight values, denoted as $W(I_{ij}) = W(C_i) + W(C_j)$, in candidate separated sets. That is, select the maximum $N(I_{ij})$ as a separated set, and if there are more than one candidate separated sets with the same maximum $N(I_{ij})$, select the one with minimum $W(I_{ij})$.
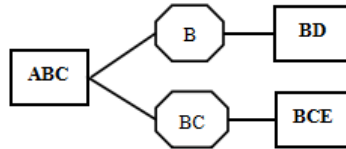
Figure.4 The local optimized junction tree

IP	
A	32%
B	34%
C	21%
D	9%
E	4%

IP Address	
A	41%
B	23%
C	19%
D	13%
E	4%

Network Layer	
A	28%
B	25%
C	32%
D	12%
E	3%

Router	
A	24%
B	26%
C	41%
D	8%
E	1%

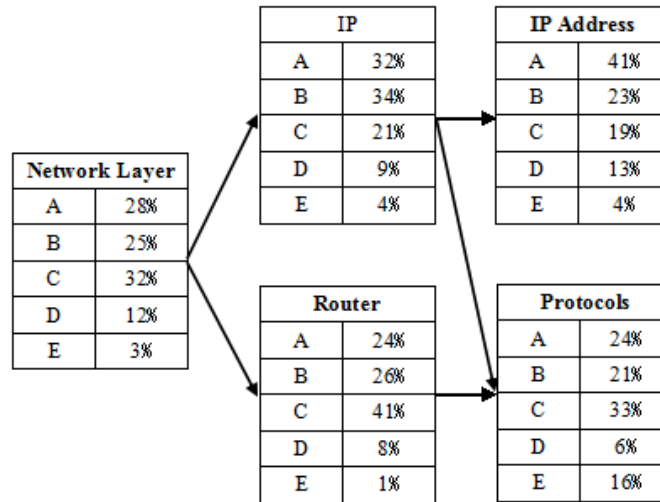| Protocols	
A	24%
B	21%
C	33%
D	6%
E	16%

Figure.5 The probability distribution after knowledge updating

In the process of creating a junction tree, in order to avoid of message reverse passing, the message from C_i to C_j can be saved in separated set, which connecting C_i and C_j. Check separated set I_{ij}, if the message exists, it has been passed, and if not, passes it. The local optimized junction tree of example for network layer belonging to course "Data Communication and Network" is shown in Fig.4. At this time, the course inference model is transformed from knowledge expression frame into inference structure of junction tree. Such structure enable to update course inference model iteratively more conveniently based on new data. As show in Fig.5, when students improve their mastery of course knowledge points, the probability distributions of other course knowledge points are updated accordingly based on message passing and course inference method.

Conclusion

In the paper, the knowledge points of network layer belonging to course "Data Communication and Network" are used as an example to show how to create a Bayesian network of course and how to transform it into a junction tree. The proposed method suggests a Bayesian network can be transformed into a local optimized tree based on observations and course historical data. In the process of transformation, inconsistent junction trees are rectified based on candidate separated sets. At last, a local optimized junction tree is created and it is used to predict local knowledge structure for students and teachers. After different local optimized junction trees are set to different weight values and every local optimized junction tree is viewed as a clique or cluster in a Bayesian network, the proposed method can be generalized to course knowledge points covering specified application field requirements, which leads up to a synthetic junction tree. In the process of course learning, collect course scores for some students by course testing, the subsequent course learning effect can be predicted based on learned course knowledge points and course inference model, so that it provide a fundamental data support to improve or enhance the following study for both students and teachers.

Acknowledgment

References

[1] Pearl J. Probabilistic reasoning in intelligent systems: networks of plausible inference[M]. Morgan Kaufmann, 2014.

[2] Almond R G, Mislevy R J, Steinberg L, et al. Bayesian networks in educational assessment[M]. Springer, 2015.

[3] Ronquist F, Teslenko M, van der Mark P, et al. MrBayes 3.2: efficient Bayesian phylogenetic inference and model choice across a large model space[J]. Systematic biology, 2012, 61(3): 539-542.

[4] Drovandi C C, Pettitt A N, Lee A. Bayesian indirect inference using a parametric auxiliary model[J]. Statistical Science, 2015, 30(1): 72-95.

[5] Vats D, Nowak R D. A junction tree framework for undirected graphical model selection[J]. The Journal of Machine Learning Research, 2014, 15(1): 147-191.

[6] Szantai T, Kovacs E. Discovering a junction tree behind a Markov network by a greedy algorithm[J]. Optimization and Engineering, 2013, 14(4): 503-518.