Abstract

An high score prior genetic simulated annealing Bayesian network structure learning algorithm (HSPGSA) by combining genetic algorithm (GA) with simulated annealing algorithm (SAA) is developed. The new algorithm provides not only with strong global search capability of GA, but also with strong local hill climb search capability of SAA. The structure with the highest score is prior selected. In the mean time, structures with lower score are also could be choice. It can avoid efficiently premature problem by higher score individual wrong direct growing population. Algorithm is applied to flight departure delays analysis in a large hub airport. Based on the flight data a BN model is created. Experiments show that parameters learning can reflect departure delay.

1. Introduction

Reducing flight delay in the air transportation system has become more urgent in recent years as air travel demand has escalated. Flight delay more precisely described as arrival and departure delay. Arrival delay, in large extent, is due to departure delay in original airport. Therefore, it focuses on the flight departure delay in this paper.

For the flight delay, scholars have conducted a number of related research. In 2003, using the method of traditional regression analysis Willy Vigneau studied flight delay and propagation[1]. Also an artificial neural network model which to be used to estimate flight departure delay was created by Dai and Liou in 2006, their paper shows that according to the input of neural network delay status may be predicted[2]. Ning Xu and his
Weidong Cao and Xiangnong Fang  /  Physics Procedia 33 (2012) 597 – 603

colleague model multi-independent airports based on Bayesian network and combined it with the model of the interaction between these airports. They discussed delay propagation between airports by Bayesian network parameters Learning[3]. In 2004, Zhengping Ma & Deguang Cui, Qinghua university, proposed an optimized model on airport flight delays. It solving the problems by Genetic algorithm and aimed at minimized the total airport delays[4]. In 2006, Lina Shi researched the airline flight delay warning evaluation, She used the multi-level fuzzy synthesis evaluation method to build the mathematical model and make the flight delays warning management[5]. For the purpose of raising flight scheduling efficiency and airspace utilization Weiwei Chen, Ri Geng & Deguang Cui suggested a heuristic algorithm and created a mixed integer programming model for the problem of arrival flight sequencing and scheduling[6].

As we all know, the civil aviation is a complex stochastic control system. Multi-uncertainties factors and the interactions between them are likely to lead to flight delays. Bayesian method is used to study multi-factors interdependence in the complex stochastic system, it combines prior knowledge and sample data to find the potential relationship in data. Bayesian network is robustness in inference and visualization. Using graph theory to build Bayesian network model to learn probability between node variables and mining uncertainty knowledge in expert system. On the one hand, causal knowledge figured by directed graph. On the other hand, statistic is expressed in conditional probability.

Combined Flight delay problem characteristic and its current research progress with advantage of Bayesian network inference and intelligent optimization algorithm, in this paper we try to incorporate genetic algorithm (GA) and the idea of simulated annealing algorithm (SA) into Bayesian network (BN) structure learning and suppos a high score prior genetic-simulated annealing approach to Bayesian network structure learning (HSPGSA). First of all, we create BN model, secondly, parameters learning is made and thirdly, flight departure delay in a large hub airport is discussed.

2. A High Score Prior Genetic-Simulated Annealing Approach to Bayesian Network Structure Learning

GA has strong global search capability, but a poor local search while SA has a strong local search and hill climber capability, but know little about whole search space. GA & SA are combined in BN structure learning based on score and search methods. The structure with the highest score is selected prior. Meanwhile the structures with lower score could also be given opportunity to be selected by improving genetic operators using SA methods. This strategy will reserve optimal gene while avoiding the premature caused by the misleading from high score individual in the population. This is the basic idea of High Score Prior Genetic-Simulated Annealing Approach to Bayesian Network Structure Learning (HSPGSA).

2.1. Algorithms flow description

For the HSPGSA, possible solutions (population) of BN structure are generated by GA. The individual (structure) with higher score is selected prior when propagation descendant population from population. Meanwhile individuals with lower score are given opportunity to be selected by SA method. This strategy will reserve optimal gene while avoiding the premature caused by the misleading from high score individual in the population.

HSPGSA algorithm flow chart see figure 1. In the algorithm initialization, the original population is generated, the individual (BN structure) score is calculated. Then, the descendant population is generated by crossover and mutation, structure individual score is calculated, the BN structure with the highest score is intended to be selected. Following is higher score prior simulated annealing module, new generation is generated. The procedure is repeated until iteration is met. The higher score prior simulated annealing module, see in figure 2, is a procedure in which new generation individuals are obtained. Individual scores of descendant population are in descending order, top n individuals are intended to be selected as a new generation individuals. Depending on annealing probabilities, individuals with lower score also could be selected.
2.2. Principal segments description

- **Scoring function**

  Scoring function is a criteria to scale BN structure individual in the population. Larger the value the better. In HSPGSA, BN structure scoring function include Maximum likelihood, Minimum Description Length and Bayesian methods. Because of its decomposable properties, BN structure score can be calculated by sum of its node score. Moreover, structure score changed when node score changed.

- **Fitness Evaluation**

  BN structure scoring function is as fitness. The process of fitness evaluation tracks the structure with higher score.

- **Crossover**

  Individuals in population paired randomly. Each paired individuals exchange part of their chromosome in crossover probabilities. Based on HSPGSA algorithms, individuals with paired exchange gene in single point method and produce a new individual.

- **Mutation**

  In order to ensure that the individuals are not all exactly the same, HSPGSA allows for a small chance of mutation. Looping through all the alleles of all the individuals in

---

![GSA_BNSL algorithm flow chart](image-url)
population, and if that allele is selected for mutation, replace it with a new value. The probability of mutation is usually between 1 and 2 tenths of a percent. Mutation is random. When combined with selection and crossover, mutation can avoid information loss and ensure genetic algorithm efficacious.

A new generation, descendant population, is generated after looping crossover and mutation through population.

- **Selection**

  During each successive generation, a proportion of the existing population is selected to breed a new generation. Individual solutions are selected through a fitness-based process, where fitter solutions (as measured by a fitness function) are typically more likely to be selected. HSPGSA sort descendant population individuals score in descending order, and then select top n (number of individuals in population) to be intend to as a new generation individuals. At the same time individuals which is not selected also have opportunity to be selected by SA. The individual can be selected in anneal probabilities. With the temperature drop, behind individuals have little chance to be selected. In this way, not only to avoid premature caused by the higher score individuals mislead, but also to make individual fitness close to the optimal solution in population. Selection operator shows in figure 2.

### 2.3 Simulation experiment

Algorithms Hill-climber (HC), SA, GA & HSPGSA are chosen separately to learning BN structure. For reliable error estimate 10-fold cross validation are used. Weka, a data mining tool, is selected as experimental environment. Initial structure of HC is Naive Bayes, Markov Blanket Classifier is used to correction factor. In SA, start temperature is set to be 10, delta temperature 0.99 and number of runs 10000. In GA, population size is 10, descendant population size 100, crossover probabilities 0.85, mutation probabilities 0.45 and number of runs 10. Because the genetic select operator is improved by SA method in HSPGSA algorithm, efficiency gains. Better BN structure can be found through genetic selection in HSPGSA than in GA. So the number of runs is halved. Annealing parameters of SA selection is same as that of SA. Statistic average is calculated for each algorithm. Experimental result is in table 1.

Among above algorithms, HC spends the least modeling time, GA has lower error but spends maximum. SA has no more advantages in learning time and accurate rate. However, HSPGSA which combine GA with SA has a better statistical average and a highest accurate rate.

### 3. Airport Flight Departure Delay Model

#### 3.1 Airport flight departure delay model based on HSPGSA

The flight departure delay Bayesian network model in a large hub airport is created by using HSPGSA, see figure 3. There are 7 nodes in it and BN graphic structure represent their causal relationship. Figure 4 shows a result of BN parameters learning. Among them, the node “航站楼” indicate flight terminal number; the node “公司” indicate Airlines; the node “任务” indicate flight task, about task and its code description is in table 2; The node “机型” indicate airplane type; the node “国际国内” indicate international or domestic flights, “I” or “D” is its value; The node “起飞时间” indicate flight departure time duration, its value “t1 to t2” represent from t1 to t2 time duration; The node “离港延误时间” indicate flight departure delay time, its value “lessthan n” represents delay time less than n minutes, “from t1 to t2” represents delay time less than t2 minutes and more than t1 minutes, “morethan n” represents delay time more than n minutes.
3.2. Parameters learning

Figure 4 is the result of BN parameters learning based on EM algorithm, it is completed in Netica analysis environment. It shows that in this airport, 82% of flights take off from No.2 terminal; The airplane type of departure flights is mainly Boeing and Airbus, which values with “B” or “A” at the beginning. Up to 47.7% of departure airplane type

![Diagram](image)

Figure 2. high score prior SA selection module

Table 1 BN structure learning algorithms comparison table

| Algorithms | statistic | Modeling time | Mean absolute error | Root squared error | Accuracy rate |
|------------|-----------|---------------|---------------------|-------------------|--------------|
| HC         | HC1       | 0.05          | 0.1482              | 0.2942            | 88.3388%     |
| HC2        | 0.06      | 0.1645        | 0.3090              | 86.5000%          |
| HC3        | 0.05      | 0.1654        | 0.3137              | 86.2000%          |
| HC4        | 0.20      | 0.2121        | 0.3241              | 86.0000%          |
| HC5        | 0.20      | 0.2273        | 0.3381              | 84.6144%          |
Table 2. Flight task code comparing table

| Task                | Code | Task                | Code |
|---------------------|------|---------------------|------|
| KB                  | SA1  | 0.11                | 0.1923|
| HB                  | SA2  | 0.1977              | 0.2950|
| General insert flight| SA3  | 0.2065              | 0.3225|
| Test flight         | SA4  | 0.2021              | 0.3192|
| Normal flight       | SA5  | 0.2166              | 0.3334|

is Boeing737. Air China (CA) flights have occupied main stream, it accounts for 38.7% of departures. China Sothern airlines (CZ) occupied 17.1%, China Eastern airlines (MU) is 14.6% and Hainan airlines (HU) is 12.4%. Also in figure 4, we have seen that domestic flights have the significantly percentage, up to 79.9% and 88.8% of tasks are normal flights in this airport.

For this large hub airport, according to relevant document issued by the Civil Aviation Administration of China, if one flight’s actual departure time later than schedule within 30 minutes, it is still normal. So in this airport, from the figure 4, majority flights are normal, it takes 57.9% of total flights. Other flights exist departure delay. Delay time of 30 minutes to 40 minutes is of 10.3%, between 40 minutes and 60 minutes is 10.76%, and so on. Few flights exist more than 2 hours delay.

Figure 3. Flight departure delay model in a large hub airport
4. Conclusions

We try to improve algorithms of Bayesian network structure learning for a higher accuracy rate. A high score prior genetic-simulated annealing approach, HSPGSA is proposed. Using a large hub airport flights data learning BN structure based on HSPGSA, comparing it with other algorithms, we have seen that the HSPGSA is able to obtain an optimized BN structure and with fast convergence rate and accuracy rate, parameters learning result can reflect airport flights departure delay.

Acknowledgements

The High Technology Research and Development Programme of China under Project 2006AA12A106 and the National Natural Science Foundation of China under Project 60879015 and the Science and Technology Programme of Civil Aviation Administration of China under Project MHRD201013 have supported this work.

References

[1] Chaug-lngHsu,Che-Chang Hsu,Hui-Chirh Li Flight Delay Propagation ,allowing for behavioural response[J].Critical Infrastructure, 2007, 3(34):301-326.
[2] Dai, D.M. and Liou, J.S. Delay Prediction Models for Departure Flights, Journal of the Transportation Research Board. CR-ROM. 2006.
[3] Ning Xu, George Donohue, Kathryn Blackmond Laskey, Chun-Hung Chen. Estimation of Delay Propagation in the National Aviation System Using Bayesian networks[J], Journal of the transportation Research Board. CR-ROM. 2007.
[4] Ma Zhengping,Cui Deguang,Airport Flight Delay Optimized Model[J], Journal of Qinghua University, 2004, 44(4), P474-477(Ch).
[5] Shi Linan,Multi-Ranks Fuzzy estimate method Applied in Flight Delay[J], Journal of Shanghai Engineering and Technology University, 2006, 20(3), P276-279(Ch).
[6] Cheng Weiwei,Geng Rui,Cui Deguang,The Optimal of Arrival Flight of Approach Area sequencing and scheduling[J], Journal of Qinghua University, 2006, 46(1), P157-160(Ch).