A New Method to Alarm Large Scale of Flights Delay Based on Machine Learning

Lu Zonglei[1], Wang Jiandong[1], Zheng Guansheng[1,2]
[1]*College of Information Science and Technology, Nanjing University of Aeronautics and Astronautics, Nanjing, 210016, China*
[2] *College of Computer and Software, Nanjing University of Information Science and Technology, Nanjing, 210044, China*
raymond@nuaa.edu.cn, jdwang@jlonline.com, guanshzh@nuaa.edu.cn

Abstract

A new method to alarm large scale of flight delays based on machine learning is presented in this paper. This new method first does unsupervised learning on the data of the flights collected from the airport. The standard of each class of delay can be gotten after the learning process. With these classes of delay, the supervised learning method can be used on the data so that the alarm model could be built. Comparing with the recent manual alarm standard, this model synthesizes more factors to do alarm. Since the recent delay standard is only related to the number of flights, which is helpful only in serious delay case, the new model performs will be more practical value than recent ones.

1. Introduction

Every year, the event "flight delay" occurs in almost all the airports. It occurs more frequently and more seriously in some busy hub-airport. Because of the serious consequences of flight delays, there are many researches about how to deal with the problem when delay occurs. However, few researches have been focused on how to alarm the delay. There are many reasons of flight delays. Macroscopically, the main reason of flight delays is that the capacity of airspace and airports cannot meet the requirement of the increasing air traffic. Bad weather, mechanical problems of aircraft, the operation issues of airlines and unexpected problems caused by passengers may be the main reasons in microscopic scales[1]. The existence of so many different factors means that the alarm could hardly been given by manual act. Thus, to use the machine learning method build the model from large history data may be the only way.

According to the general machine learning methods, there should be a special attributes named "class" in the data set in order to train the alarm model. However, the recent standard to classify the level of flight delays is so subjective that some level may never be reached. For example, there are four levels of delays in the standard of some hub-airport of China in the following.

Case 1, Blue Level: there should be a blue alarm if more than 40% of the departure flights will be delayed for the bad weather or the control of route.

Case 2, Yellow Level: there should be a yellow alarm if the weather is so bad that more than 60% of the departure flights in the coming two hours will be delayed or more than 10 flights will be delayed for more than 4 hours.

Case 3, Orange Level: there should be an orange alarm if more than 80% of the departure flights in the coming two hours will be delayed for the bad weather.

Case 4, Red Level: there should be a red alarm if all of the departure flights in the coming two hours will be delayed for the bad weather.

Only weather factor has been contained in this standard. However it is not the main factor of flight delays according to paper [1]. There are only 4.63% of delayed flights caused by the bad weather. Therefore, it is difficult to check the validity of this standard.

2. Unsupervised learning model

Usually, the number of delayed flights, the number of delayed passengers, the airport which the delays have spread to, the number of air companies whose flights has delayed and the total delay time are the common indexes to grade the flight delays. Any neglected indexes may invalidate the delay grades. From the viewpoint of data, there should be some similarity of the data in the same delay grade on the above indexes. Thus, we can get the delay grades (or classes) by group the flight data with the similarity, which is just a process of clustering. Clustering is perceived as an unsupervised process since there are no predefined classes and no examples that would show what kind of desirable relations should be valid among the data[2]. Clustering is a process of grouping the data into classes or clusters so that data within a cluster exhibit high similarity, but are very dissimilar to data in other clusters[3]. An important problem of clustering applications is to find the optimal (or local optimal) parameters, which requires the clustering validity indexes.

978-0-7695-3488-6/08 $25.00 © 2008 IEEE
DOI 10.1109/KAM.2008.18

To simplify the problem, we will use *k*-Means algorithm to do clustering, whose parameter is just the number of clusters. The data are the real records of flights in the hub-airport of China in 2006. The data have been standardized in order to offset the difference of absolute values of different attributes. The number of clusters varies from 3 to 10. Each clustering has been measured by Dunn's Index[4], Davies-Bouldin's Index[5], Classification Entropy[6], CS Index[7], Separation Index[8] in order to find the optimal parameter. Table1 table shows the experiments.

Table 1 The Indexes Values of Each Clustering

Number of Clusters	Dunn's Index	Davies-Bouldin's Index	CS Index	Classification Entropy	Separation Index
3	0.0417	0.7622	1.3653	0.3804	0.1870
4	0.0373	0.7944	1.3290	0.4860	0.2150
5	0.0454	0.8357	1.2638	0.5766	0.2483
6	0.0353	0.9170	1.4038	0.6453	0.3165
7	0.0372	0.9182	1.2662	0.7103	0.4458
8	0.0353	0.9435	1.3217	0.7713	0.3610
9	0.0376	1.0129	1.3776	0.8252	0.4188
10	0.0305	0.9918	1.3257	0.8729	0.4236
Optimal Number	5	3	5	3	3

From Table1, we can get that some indexes do not agree with others to choose the optimal number. This may be caused by the unconventional similarity has been used in the clustering of flights data. To avoid the vagueness of different similarity, we need a universal clustering index which could handle unconventional similarity. Lu's Index presented in Paper [9] is a universal clustering index based on modal logic, which could be used in the case of different similarity. Table2 show the result of Lu's Index of the above clustering.

Table 2 Lu's Indexes Value of Each Clustering

Number of Clusters	3	4	5	6	7	8	9	10	Optimal Number
Lu's Index	0.1884	0.1719	0.1704	0.2006	0.2083	0.2318	0.2514	0.2261	5

We can get that there should be five grades of flights delay in this airport from Table1 and Table2. If these grades contain the one of undelayed flights, the number will agree with the manual standard of this airport. Table3 shows the information of these five classes.

Table 3 Information of Each Class

Grades	Delayed Flights	Delayed Time	Spread Airports	Spread Company	Delayed Passengers
Red	0.8233±0.0934	0.3867±0.174	0.8299±0.0757	0.7859±0.1246	0.8276±0.0921
Orange	0.5615±0.0685	0.1539±0.0515	0.6784±0.0588	0.6778±0.0997	0.5774±0.0649
Yellow	0.3820±0.0585	0.0885±0.0217	0.5333±0.0522	0.5579±0.0938	0.3991±0.0575
Blue	0.2342±0.0508	0.0578±0.0185	0.3893±0.0574	0.4130±0.0898	0.2436±0.0531
White	0.1113±0.0576	0.0367±0.023	0.2094±0.0967	0.2084±0.0961	0.1170±0.0605

3. Alarm Model

The unsupervised learning model gives the class of each record. With these classes, we can train the classification models as the alarm model. The so-called "alarm model" is just a group of judgment rules, by which we can predict the delay in the future. If we want to given an alarm class in some time, for example at 8:00am, we need to build a training data set of the number of delayed flights, the number of delayed passengers, the airport which the delays have spread to, the number of air companies whose flights has delayed and the total delay time in each day before the given time. And then we can train the classification model with this set.

The common classification methods include Bayes models, decision trees, neural networks and rule models. The different model plays well in different application conditions. There are so few recent researches about the flight data that we cannot judge which model will be better before the experiments. The only way is to train the models and compare the statistical results of each model. We try to train the naive Bayes model[10], C4.5 decision tree model[11], BP neural network model and Ridor rule model[12] by the standardized data of each day before 8:00 in 2006 of the hub-airport in China. Table4 shows the statistical results of each model.

Table4 Statistical Results of Each Model

Models	Incorrectly Classified Instances	Kappa statistic	Mean absolute error	Root mean squared error	Relative absolute error	Root relative squared error
Naive Bayes	54.5205%	0.2967	0.2266	0.3966	75.0904%	102.1397%
C4.5 Decision Tree	20.274%	0.7291	0.1212	0.2462	40.1601%	63.3979%
BP Neural Network	38.3562%	0.4771	0.2075	0.3214	68.7521%	82.7804%

Ridor	36.1644%	0.5263	0.1447	0.3803	47.9322%	97.9503%

From the table, we can get that C4.5 decision tree model performs best. The tree is shown in Figure1. From the figure, we can get that the number of delayed flights is not as important as the manual standard. Under the node "flights", there are no white nodes and only a few blue nodes. It means that this index only can help to decide whether there will be serious delays.

Considering all the five indexes comprehensively, the model shown in Figure1 gives the alarm rules of 8:00. Similarly, we can get other rules of 9:00, 10:00 or any other time. Then we can predict the class of delay. From Table4, we can get that the confidence of the prediction will be no less than 80% so that both airport and air companies could make the decision with this prediction.

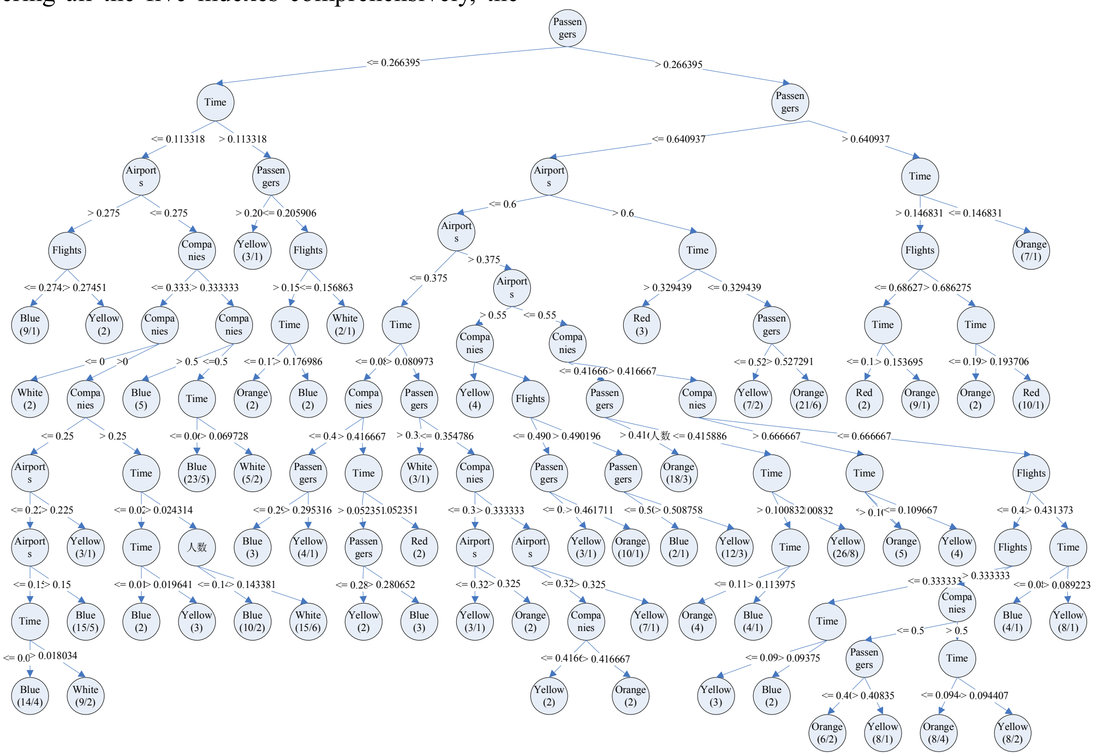

Figure 1 C4.5 Decision Tree Model Training by Flights Data

5. Conclusions

Alarming the flights delays is a world-class difficult problem since there are so many factors may cause delay. We have presented a new method to discover the delay grades from the flight data. With the grades, some alarming models have been trained. The best one has been chosen as the final model. The experiments show that the confidence of this model is no less than 80%. Thus, it will be helpful to alarm the coming delays.

Acknowledgement

This project is supported by 863-Hi-Tech Research and Development Program of China (2006AA12A106).

References

[1] Ma Zhengping, Cui Deguang. Optimizing airport flight delays. Journal of Tsinghua University (Science and Technology), 2004, 44(4): 474-477,484.

[2] M. Halkidi, Y. Batistakis and M. Vazirgiannis: Clustering validity checking methods: part II, SIGMOD Record, 2002, 31(3): 19-27

[3] Lu Zonglei, Wang Jiandong. Fast Clustering Algorithm Based on Hypersphere of Multidimensional Space. Journal of Nanjing University of Aeronautics and Astronautics, 2006, 38(6): 706-711

[4] Dunn, J.C. Well separated clusters and optimal fuzzy partitions, Journal of Cybernetics, 1974, 4(1): 95-104.

[5] Davies DL, Bouldin DW. A cluster separation measure. IEEE Trans on Pattern Analysis and Machine Intelligence, 1979, 1(4): 224-227

[6] Bezdek JC. Numerical taxonomy with fuzzy set. Journal of Mathematical Bilolgy, 1974, 1(1): 57-71

[7] Chou C.H, Su M.C, Lai. E. A new cluster validity measure and its application to image compression. Pattern Analysis and Applications, 2004, 7(2): 205-220.

[8] Xie XL, Beni G. A validity measure for fuzzy clustering[J]. IEEE Trans on Pattern Analysis and Machine Intelligence, 1991, 13(4): 841-847

[9] Lu Zonglei, Wang Jiandong. An index of cluster validity based on modal logic. Journal of Computer Research and Development. 2008, 45(9): 1477-1485
[10] George H. John, Pat Langley: Estimating Continuous Distributions in Bayesian Classifiers. In: Eleventh Conference on Uncertainty in Artificial Intelligence, San Mateo, 1995, 338-345.
[11] Quinlan, J. R. C4.5: Programs for Machine Learning. Morgan Kaufmann Publishers, 1993.
[12] Brian R. Gaines, Paul Compton. Induction of Ripple-Down Rules Applied to Modeling Large Databases. Journal of Intelligent Information Systems,1995, 5(3): 211-228.