13

Alarming Large Scale of Flight Delays: an Application of Machine Learning

Zonglei Lu
College of Computer Science and Technology,
Civil Aviation University of China
People's Republic of China

1. Introduction

Flight delays occur at almost all the airports every year. At major hub airports, flight delays occur more frequently and cause more serious problems. The FAA, i.e. Federal Aviation Administration, divides the flight delays into two types. One is the so-called technique delay, which is mainly caused by the flight waiting for the resources of air flow control or the limit of flow management, and the other is the so-called effective arrival delay. According to some papers, there are five kinds of flight delays, i.e. the delay of flow operating, the delay of aircraft ground operating, the aircraft technique delay, the delay of flow control and the airport 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. In recent years, most of the researchers agree with that the main reason of flight delays is the limit of the runway capacity. Therefore flight delays can be reduced by increasing the number of runways, taxiways and the air flow control devices and improving flight management, such as the process of air flow control, the sequence of the arrival and departure flights.

Because of the serious consequences of flight delays, there has been a tremendous amount of researches on how to deal with the various problems caused by flight delays. However, very little research focuses on how to forecast flight delays. Since there are so many factors of the flight delays, it is very difficult to calculate the delays by the deduction. Any mistakes of the factors may lead the result invalid; in contrast, the prediction by statistical law may be more effective. From the view of statistics, the flights could be perceived as repeated experiments, although there may be a few differences between the environments of each experiment. Therefore, the delay time of the flight should satisfy some special probability distribution. The difference of the experiment environments is just the reason why the time taken by the flights is variable. The uncertainty of experiment environments would add the prediction difficulty of delay time. Conversely, assume that all the factors of two flights are same, i.e. the two experiments have been taken in the same environment. Then they must have same delay time, i.e. the results of the two experiments are same. The probability distribution

could show the a priori probability of the delay time. As a prediction system, the posterior probability, i.e. the probability of the delay time in some special condition, may be more important than the a priori probability. Machine learning is usually used to calculate the posterior probability from the data.

In this chapter, an example will be shown to introduce how to use machine learning methods to alarm large scale of flight delays, which is a very complex problem. This is a real application in a hub airport of China. We will use the symbols "*Airport A*" to denote this airport with omitting its real name. The concept of "flight delay" may be changed under different requirement. Thus, the background of the application will be introduced firstly in this chapter and the relative concepts, especially the application requirement, will be defined clearly. Besides, the data gotten from the practice, which is an important factor to use machine learning methods, will be depicted in this chapter. According to the data and the requirement, the process of learning could be discussed. To solve the problem of alarming flight delays, the delay levels, which are used to describe the serious degree of delays, are required. Therefore, there will be a section to show how to use the unsupervised learning method to calculate the delay levels. Depending on the defined levels, the supervised method could be used to predict the coming delays. Some classical methods, both supervised and unsupervised, will be tried in this chapter, although there will be only one method in the final application. After introducing these methods, how to choose the most suitable one will be discussed. At last, a conclusion of the whole chapter will be shown .

2. Application Requirements and Data Preparing

Flight is a civil aircraft which take a mission from a certain departure airport to the certain arrival airport in the certain time according to the flight schedule. There are two kinds of time to each flight in the schedule, i.e. the plan departure time and the plan arrival time. The flight departure or arrival at the time later than the plan departure or arrival time is called a departure delay flight or arrival delay flight respectively. Since the flight is influenced by multiple factors, the concepts of departure delay flight and arrival delay flight may be too strict. The constraint is usually relaxed in practical applications. In the *Airport A*, the flight which departs or arrives at the time later than the schedule time in 30 minutes is treated as a normal flight but not a delayed one. The concept "large scale of flight delays" means the number (or the ratio) of the delayed flights is more than a given threshold. To alarm large scale of flight delays is just to give a prediction, the main foundation of which is the recent status of the airport, before the delays.

The main difference between engineering application and laboratory experiment is the data, which are always ready in the laboratory. Collecting enough data and converting them into right form is very important to the engineering. The statistics theory, which aims at the asymptotic theory when sample size is tend to infinity, is the basic theory of machine learning. However, the size of sample usually is finite, so that some learning methods perform well only in theory. To guarantee the application of the machine learning method effectively, we need discuss the form of the data which could be collected from the airport. The form of data restricts the method which could be used in the application. Conversely, the method itself could tell us what kind of data are required.

Intuitively, finding the association event of flight delays may be a good way to do prediction. The alarm could be given when the association events happened. The recent researches show that there are five classes of factor which may lead flight delays, including: weather, the management of the airline company, the control of airports, the air traffic control and the factor of the passengers. However, it is not so easy to collecting the data of all these factors. For example, the factor of the passengers is really uncertainty. It is possible that some passenger, who had already passed the security checking, decided not to board in without any announcement. In this case, the staffs must find the passenger and confirm that he (or she) would not board in. The flight could not take off before the confirmation. Since this is just a stochastic event which may cause the flight delays, it is very difficult to monitor this factor. Therefore, the prediction from the factors may be invalidity.

Actually, the so-called "association event" may not be the direct reason of delay, but a concept of statistics. An event is called an association event of the flight delays means that the flight is very likely to delay when this event occurs. In the language of the probability theory, the conditional probability of the association event to the flight delays is large enough. Thus, we need to find the association events of the flight delays firstly, and then to find the pattern of the association events and the flight delays. Obviously, only the events in the records of *Airport A* could be used to predict flight delays. Therefore, we will show the Attributes of the records of *Airport A* in the following (See Table 1) in order to discuss the association events.

Attribute	Explanation	Type
FID	The ID of the flight	Nominal
AID	The ID of the aircraft	Nominal
D/A	Departure flight or arrival flight where the symbol "D" and "A" stands for departure flight and the arrival flight, respectively	{"D", "A"}
D/I	Domestic flight or International flight where the symbol "D" and "A" stands for Domestic flight or International flight, respectively	{"D", "I"}
AT	The type of the aircraft	Nominal
CT	The code of the task such as normal flight, charter flight and so on	Nominal
TB	Terminal building which will serves the flight	Numeric
AC	The air company that the flight belongs to	Nominal
SD	The time when the flight departs in the schedule	Numeric
PD	The predicting time when the flight will depart	Numeric
RD	The time when the flight departs	Numeric
TA	The target airport of the flight	Nominal
SA	The time when the flight arrives in the schedule	Numeric
PA	The predicting time when the flight will arrive	Numeric
RA	The time when the flight arrives	Numeric
BS	The boarding status of the flight	Nominal
BG	Which gate will be used to boarding to the flight	Nominal

Table 1. the Attributes of the records of *Airport A*

Table 1 shows all attributes recorded by *Airport A*. In other words, all predictions must be done only on these attributes. However, there is very little correlation between the above

attributes and the delays intuitively. Actually, the calculation of the data shows the same conclusion, i.e. none of the above attributes is the association event of flight delays. This may be a big problem. Without the association events, the prediction will miss the basis.

The above case is very common in many applications. Data are not so prefect where we cannot find what we expected. In this case, we must rebuild the data set with the association attributes. Rebuilding data set does not mean recollecting data. What we need to do is just to get the statistical information, which is associated to the flight delay intuitively.

One of the common methods is to build time associated attributes, i.e. the association of recent status and the past status. In most of systems, the recent status is an important factor of the future status. Since the computer could only process the discrete data, we assume that the monitoring system is discrete. For example, we can assume that the status is monitored every five minutes. Because the flight delay is mainly caused by the capacity of airspace and airports not large enough, only a few flights could be executed in five minutes. Thus, the recent delay flights in the waiting sequence may be also in the waiting sequence in the next five minutes. The number of delay flight in the next statistical period includes two parts: one is the recent delay flights and the other is the new arrival or departure delay flights. In other words, the recent status is associated with the future status. Furthermore, we can use the association to do prediction in order to alarm the flight delays.

The severity of flight delays finds expression not only in the amount of delayed flights, but also in the delay hours and passenger numbers involved, etc. Since we are going to build a model to predict the future delays, we must try to find all association events as possible as we can. According to the recent researches, there five aspects of flight delays including the number of delayed flights, total delay hours, the number of passengers involved, the number of air companies involved, and the number of airports which the delays have spread to directly. We can get all the value of the above attributes by statistical method except the number of passengers involved. The data provided by *Airport A* does not contain the information of passengers. But we can get an estimate value by the number of seats of each type of aircraft and the average attendance rate, which could be found in the civil aviation handbooks. Table 2 shows some samples of the association attributes in the following.

the number of delayed flights	total delay hours	the number of airports which the delays have spread to directly	the number of air companies involved	the number of passengers involved
118	142	61	24	14997
47	79	32	14	5904
85	301	49	22	10318
110	144	56	26	14160
156	147	65	32	20819
169	243	68	34	22411
172	219	62	27	22951
150	137	59	30	20074
102	115	55	21	13588
173	206	64	23	22140
…	…	…	…	…

Table 2. Samples of the Association Attributes to Flight Delays

The data shown in Table 2 are statistical values of each day at a given time. Considering there are big differences between the ranges of each attribute shown in Table 2. The attributes with the big range may influence the prediction more than the one with small range, which is unfair intuitively. We will normalize the data in order to eliminate effects.

3. Grading Flight Delays with Unsupervised Learning Methods

According to the data shown in Table 2, our prediction should be aim at the delay levels of *Airport A* at a given time. The basis of the prediction is the value of the five attributes shown in Table 2. Thus, we must build a model to show the relationship of these five attributes and delay levels. However, there is no information about the delay levels. Therefore, we must calculate the delay levels at given time according to the value of the five attributes shown in Table 2. The prediction model could be built only after the levels have been ready.

From the viewpoint of data, there should be some similarity among the data in same level. Assume that there is only one attribute (for example, the number of delayed flights) in Table 2. Then the values of data in the same level must be very close to each other. Similar with this simple case, the data in the same level must be similar with each other in the case of more than one attributes. The similarity could be calculated by the sum of squares of the difference of each attribute, which denotes the Euclid's distance, between two data. Thus, the level could be gotten 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.

To simplify the problem, we will use *k*-means algorithm, which is very popular and easy to be implemented, to grade the data of flight delays. There are three steps of *k*-means algoithm. First of all, select *k* means of clusters randomly. Then put each data to the cluster whose mean is nearest to the data, i.e. the similarity measure value of the data and the cluster's mean is bigger than the value of any other cluster's mean and the data. Then the algorithm recalculates the mean of each cluster. At last the algorithm output the clusters if they are not changed, otherwise it will return to the second step. The parameter *k* denotes the number of clusters, which means the number of levels of flight delays in this application. The recent subjective standard of flight delays in *Airport A* divides the flight delays into four levels, which is shown in Table 3.

Levels	Definition
Blue Level	there should be blue alarm if more than 40% of the departure flights will be delayed for the bad weather or the control of route
Yellow Level	there should be 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.
Orange Level	there should be orange alarm if more than 80% of the departure flights in the coming two hours will be delayed for the bad weather.
Red Level	there should be red alarm if all of the departure flights in the coming two hours will be delayed for the bad weather.

Table 3. Recent Standard of Flight Delays in *Airport A*

The main factor of the levels shown in Table 3 is the weather, but there are only about 4.63% of delayed flights caused by the bad weather according to (Ma & Cui, 2004). Therefore, it is difficult to check the validity of this standard. Besides, there is no attribute about the weather in the data set provided by *Airport A*. Thus, we cannot mark the levels to the data.
Although this subjective grading may not fit the data very well, it still provides some information about the level. After all, these levels are constituted by the the domain experts. There are four levels in Table 3 without the level of few flight delays. Thus, it may be five levels of the data, which is is an important information of the *k*-means algorithm. According to Table 3, the parameter of the *k*-means algorithm may be set as *k*=5 in this application. Certainly, this is just an assumption which needs to be checked in the experiment.
Since there is no standard answer about what the clustering process learns, a measure index of clustering is required in order to make the clustering result fit the application and find the optimum parameters. We first introduce four clustering validity indexes, including Dunn's index, Davies-Bouldin's index, CS index and Lu's index, and then try to find the optimum parameter *k* of *k*-means by these four indexes.
Dunn's index is defined as

$$D_{n_c} = \min_{i=1,\ldots,n_c} \left\{ \min_{j=i+1,\ldots,n_c} \left\{ \frac{d\left(C_i, C_j\right)}{\max diam(C_k)} \right\} \right\} \tag{1}$$

where $d\left(C_i, C_j\right)$ is the dissimilarity function between two clusters C_i and C_j defined as $d\left(C_i, C_j\right) = \min_{x \in C_i, y \in C_j} d\left(x, y\right)$ and *diam(C)* is the diameter of a cluster, which may be considered as a measure of clusters' dispersion. The diameter of a cluster *C* can be defined as follows:

$$diam\left(C\right) = \max_{x, y \in C} d\left(x, y\right) \tag{2}$$

If the data set contains compact and well-separated clusters, the distance between the clusters is expected to be large and the diameter of the clusters is expected to be small.
The Davies-Bouldin's index is defined as

$$DB_{n_c} = \frac{1}{n_i} \sum_{i=1}^{n_c} R_i \tag{3}$$

where $R_i = \max_{j=1,\ldots,n_c, i \neq j} \frac{s_i + s_j}{d_{ij}}$ and s_i is a measure of dispersion of a cluster C_i. It is clear for the above definition that DB_{n_c} is the average similarity between each cluster C_i, i=1, ..., n_c and its most similar one. It is desirable for the clusters to have the minimum possible similarity to each other; therefore we seek clusterings that minimize *DB*. The DB_{n_c} index exhibits no trends with respect to the number of clusters and thus we seek the minimum value of DB_{n_c} in its plot versus the number of clusters.
The CS index is defined as

$$CS(N_c)\frac{\sum_{i=1}^{N_c}\left\{\frac{1}{|C_i|}\sum_{x_j\in C_i}\max_{x_k\in C_i}\left\{d(x_j,x_k)\right\}\right\}}{\sum_{i=1}^{N_c}\left\{\min_{j=1,2,..,N_c,j\neq i}\left\{d(v_i,v_j)\right\}\right\}} \tag{4}$$

where x_j and v_i is the jth data and the center of the ith cluster, respecitvely.
The Lu's index is defined as

$$LI(N_c)=\frac{1}{2}\sum_{i=1}^{N_c}\sum_{x_j\in C_i}\left(\left|\left\{d_k\in C_j : d(d_j,d_k)>\lambda\right\}\right|+\left|\left\{d_k\notin C_j : d(d_j,d_k)\leq\lambda\right\}\right|\right) \tag{5}$$

where λ is a threshold of similarity. The Lu's index is not with respect to the concept of "cluster center" so that it could be used in the case of non-spherical clusters.
We will cluster the data by k-means algorithm with the parameter k varies from 3 to 10. The value of the above indexes on the clusters is shown in Table 4.

Number of Clusters	Dunn's Index	Davies-Bouldin's Index	CS Index	Lu's Index
3	0.0417	0.7622	1.3653	0.1884
4	0.0373	0.7944	1.3290	0.1719
5	0.0454	0.8357	1.2638	0.1704
6	0.0353	0.9170	1.4038	0.2006
7	0.0372	0.9182	1.2662	0.2083
8	0.0353	0.9435	1.3217	0.2318
9	0.0376	1.0129	1.3776	0.2514
10	0.0305	0.9918	1.3257	0.2261
Optimal Number	5	3	5	5

Table 4. The Values of Indexes on Each Clustering

Level	the number of delayed flights	total delay hours	the number of airports which the delays have spread to directly	the number of air companies involved	the number of passengers involved
very low delays	0.1113±0.0576	0.0367±0.023	0.2094±0.0967	0.2084±0.0961	0.1170±0.0605
low delays	0.2342±0.0508	0.0578±0.0185	0.3893±0.0574	0.4130±0.0898	0.2436±0.0531
moderate delays	0.3820±0.0585	0.0885±0.0217	0.5333±0.0522	0.5579±0.0938	0.3991±0.0575
significant delays	0.5615±0.0685	0.1539±0.0515	0.6784±0.0588	0.6778±0.0997	0.5774±0.0649
excessive delays	0.8233±0.0934	0.3867±0.174	0.8299±0.0757	0.7859±0.1246	0.8276±0.0921

Table 5. The Mean and Standard Deviation of Each Cluster

According Table 4, there should be five levels of flight delays, which is consistent with the conclusion of the domain experts. Intuitively, these five levels should be very low delays, low delays, moderate delays, significant delays, excessive delays, respectively. The mean and the standard deviation of each cluster is shown in Table 5.
For each formula x±y in the grids of Table 5, x and y denotes the mean and standard deviation, respectively. As a learning result from data, these levels may fit the data better than the recent subjective standard.

4. Train the Supervised Learning Methods

The unsupervised learning model builds the levels of the flight delays, which could act as the attribute "class" of the records. With this attribute, we can train the classification models as the alarmming model. Notice that it is not always validity to training classification models with the attribute built by clustering. The clustering is a process to build the classes (clusters) by the given relation (similarity) of the data while the classification is to find the relation of the data by the given classes. Thus, the clustering and the classification are the inverse processes of each other in some sense. The relation found by the classification does just fit to the class. If the class is built by clustering, the one which relation fits to is just the relation on which the clustering is based. Then the relation learning by the classification on the same data set may be invalid. In the application of alarming large scales of flight delays, the supervised learning method is used to find the relationship between the recent status and the future delays, which is unknown before training. Therefore, the unsupervised method and the supervised method in this application are not on the same training data set.
According to the theory of machine learning, the result of the learning method fits both the a priori knowledge and the data. Therefore, the a priori knowledge is an important factor of the effective of learning result. Usually, the a priori knowledge of the supervised learning is the structure of the model. Thus, we must choose a suitable model in order to meet with good results. The decision tree model, Bayesian learning model and the artificial neural networks model are the most common classification methods. We will try to build the alarming model based on these methods. Before the application, there is a brief introduction of these method in the following.
The problem of constructing a decision tree can be expressed recursively. First, select an attribute to place at the root node and make one branch for each possible value. This splits up the example set into subsets, one for every value of the attribute. Now the process can be repeated recursively for each branch, using only those instances that actually reach the branch. If at any time all instances at a node have the same classification, stop developing that part of the tree.
C4.5 builds decision trees from a set of training data, using the concept of information entropy. At each node of the tree, C4.5 chooses one attribute of the data that most effectively splits its set of samples into subsets enriched in one class or the other. Its criterion is the normalized information gain (difference in entropy) that results from choosing an attribute for splitting the data. The attribute with the highest normalized information gain is chosen to make the decision. The C4.5 algorithm then recurses on the smaller sublists.
The following shows the main steps of C4.5. First of all, for each attribute *a*, find the normalized information gain from splitting on *a*. And then let *a_best* be the attribute with the highest normalized information gain. Create a decision node that splits on *a_best*. At last,

recurse on the sublists obtained by splitting on *a_best*, and add those nodes as children of node.

One highly practical Bayesian learning method is the naive Bayes learner, often called the naive Bayes classifier. The naive Bayes classifier applies to learning tasks where each instance x is described by a conjunction of attribute values and where the target function $f(x)$ can take on any value from some finite set V. A set of training examples of the target function is provided, and a new instance is presented, described by the tuple of attribute values ($a_1, a_2, ..., a_n$). The learner is asked to predict the target value, or classification, for this new instance.

The naive Bayes classifier is based on the simplifying assumption that the attribute values are conditionally independent given the target value. In other words, the assumption is that given the target value of the instance, the probability of observing the conjunction $a_1, a_2, ..., a_n$ is just the product of the probabilities for the individual attributes.

The study of artificial neural networks has been inspired in part by the observation that biological learning systems are built of very complex webs of interconnected neurons. In rough analogy, artificial neural networks are buit out of a densely interconnected set of simple units, where each unit takes a number of real-valued inputs (possibly the outputs of other units) and produces a single real-valued output (which may become the input to many other units)

Backpropagation, or propagation of error, is a common method of teaching artificial neural networks how to perform a given task. It is a supervised learning method, and is an implementation of the Delta rule. It requires a teacher that knows, or can calculate, the desired output for any given input. It is most useful for feed-forward networks (networks that have no feedback, or simply, that have no connections that loop). The term is an abbreviation for "backwards propagation of errors".

The backpropagation neural network works in the following steps. Present a training sample to the neural network. Compare the network's output to the desired output from that sample. Calculate the error in each output neuron. For each neuron, calculate what the output should have been, and a scaling factor, how much lower or higher the output must be adjusted to match the desired output. This is the local error. Adjust the weights of each neuron to lower the local error. Assign "blame" for the local error to neurons at the previous level, giving greater responsibility to neurons connected by stronger weights. Repeat from above step of calculating what the output should have been on the neurons at the previous level, using each one's "blame" as its error.

Usually, the common method to choose models is to analyze the data, especially the intuitive meaning of the attributes. However, 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. The Kappa statistic is a common statistical measure of inter-rater agreement for qualitative (categorical) items. It is generally thought to be a more robust measure than simple percent agreement calculation since Kappa statistic takes into account the agreement occurring by chance. The Equation 6 shows the formal definition of Kappa statistic in the following.

$$K = \frac{\Pr(a) - \Pr(e)}{1 - \Pr(e)} \tag{6}$$

where $\Pr(a)$ is the relative observed agreement among raters, and $\Pr(e)$ is the hypothetical probability of chance agreement, using the observed data to calculate the probabilities of each observer randomly saying each category. If the raters are in complete agreement then $K = 1$. If there is no agreement among the raters (other than what would be expected by chance) then $K \le 0$.

Besides, the mean absolute error, the root mean squared error, the relative absolute error and the root relative squared error are other common indexes to measure the learning result. These indexes could be easily found in the statistical books so that we do not show the formulas here.

We will try to train the above three supervised learning models on the data with the classes built by the unsupervised method, and then choose the best one. The training data set is built based on the normalized data of each day after 8:00 in some year of *Airport A*. Table 6 shows the 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%

Table 6 Statistical Results of Supervised Learning Models

As Table 6 shows, the C4.5 decision tree model is the best one according to the statistical values. Thus, we can use this model to predict the fight delays. It may be the final step of laboratory experiment that the model has been trained ready. However, this is an engineering application, but not a laboratory experiment. In the engineering application, the intuitive meaning of the model is required in order to check the validity by experiences. The decision tree is not intuitive enough to be read. A common way to solve this problem is to convert the decision tree to rules, which is very easy to read. And then the rules may be modified by the domain experts in the applications. The finally model to alarm large scale of flight delays is built by these checked and modified rules.

Since the alert of large scale of flight delay is a momentous decision, the alarming model must be controllable and explainable. Good statistical result is only a necessary condition. This is very important in the engineering applications, but usually not required in the laboratory. Compare with the Bayesian models and the artificial neural network models, the decision tree models could be explained more intuitively. Then the users could check whether it needs to be modified by their domain knowledge and experience. Thus, the model with clearly intuitive meaning is always chosen in the applications.

5. Conclusion

This chapter shows an example to use machine learning method in applications. Limited by the space of pages, some details have been skipped to get the main point. As a supplement, we will show some reference in which the details could be found. The flight delays will cause a series of serious consequences. However, it is very difficult to predict flight delays accurately. Some models have been presented to describe the flight delays, such as Milan Janic's disruption model (Janic, 2005), Ning Xu's Bayesian network model (Xu, 2007) and Zonglei Lu's decision tree model (Lu et al, 2008a) and so on. However, there is no model could predict the flight delays accurately up to now. These models could give only some reference of the prediction.

There are some machine learning methods mentioned in this chapter. The systematic introduction about these methods could be found in the classical machine learning books, such as (Mitchell, 1997) and (Witten & Frank, 2005). For more details, the *k*-means clustering algorithm is first mentioned in (Jain, 1967). The Dunn's index, Davies-Bouldin's index, CS index and Lu's index has been introduced by (Dunn, 1974), (Davies & Bouldin, 1979), (Chou et al, 2004) and (Lu et al, 2008b), respectively. Some details about these clustering validity indexes could also be found in (Halkidi et al, 2002), a survey of clustering validity index.

The C4.5 decision tree model, the naive Bayes model and the BP neural network model is mentioned in (Quinlan, 1993), (John & Langley, 1995), (Werbos, 1974), respectively.

6. References

Chou C.H; Su M.C & Lai. E. (2004). A new cluster validity measure and its application to image compression. *Pattern Analysis and Applications*, Vol. 7, No. 2. (June 2004) pp.205-220, ISSN: 1433-7541.

Davies D.L. & Bouldin D.W. (1979). A cluster separation measure. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 1, No. 2 (April 1979) pp.224-227, ISSN: 0162-8828

Dunn, J.C. (1974). Well Separated Clusters and Optimal Fuzzy Partitions. *Journal of Cybernetica*, 1974, Vol. 4, No. 1 (January 1974) pp.95-104. ISSN: 0022-0280

Halkidi M.; Batistakis Y. & Vazirgiannis M. (2002). Clustering validity checking methods: part II. *SIGMOD Record*, Vol. 31, No. 3 (September 2002) pp.19-27

Janic, M. Modeling the Large Scale Disruptions of an Airline Network. *Journal of Transportation Engineering*. Vol. 131, No. 4 (April 2005) pp.249-260. ISSN: 0733-947X

Jone, G.H. & Langley, P. (1995). Estimating Continuous Distributions in Bayesian Classifiers. *Proceedings of the Eleventh Conference on of Uncertainty in Artificial Intelligence*. pp. 338-345, ISBN: 9781558603851, Montreal CA, Morgan Kaufmann, San Francisco, USA

Lu Z.L.; Wang J.D. & Li Y. (2008a). An index of cluster validity based on modal logic. *Journal of Computer Research and Development*. Vol. 45 No. 9. (September 2008) pp.1477-1485. ISSN: 1000-1239

Lu Z.L.; Wang J.D. & Zheng G.S. (2008b). A New Method to Alarm Large Scale of Flights Delay Based on Machine Learning. *Proceeding of 2008 International Symposium on Knowledge Acquisition and Modeling*, pp.589-592, ISBN: 9780769534886, Wuhan China, (December 2008), IEEE CS, New York US.

Ma Z.P. & Cui D.G. (2004). Optimizing airport flight delays. *Journal of Tsinghua University (Science and Technology)*, Vol.44 No.4, (April 2004) pp.474-477,484. ISBN: 1000-0054

Mitchell T.M. (1997) *Machine Learning*, McGraw-Hill, ISBN: 0070428077, New York US.

Quinlan, R. (1993). *C4.5: Programs for Machine Learning*. Morgan Kaufmann, ISBN: 1558602380, San Mateo, CA.

Werbos, P.J. (1974). Beyond Regression: New Tools for Prediction and Analysis in the Behavioral Sciences. PhD thesis, Harvard University.

Witten, I.H. & Frank, E. (2003). *Data Mining: Practical Machine Learning Tools and Techniques Second Edition*. Morgan Kaufmann, ISBN: 9780120884070

Xu, N.; Laskey, K.B.; Chen, C.H. et al (2007). Bayesian Network Analysis of Flight Delays. *Proceeding of Transportation Research Board 86th Annual Meeting*. Washington DC, US.

Machine Learning

Edited by Yagang Zhang

ISBN 978-953-307-033-9
Hard cover, 438 pages
Publisher InTech
Published online 01, February, 2010
Published in print edition February, 2010

Machine learning techniques have the potential of alleviating the complexity of knowledge acquisition. This book presents today's state and development tendencies of machine learning. It is a multi-author book. Taking into account the large amount of knowledge about machine learning and practice presented in the book, it is divided into three major parts: Introduction, Machine Learning Theory and Applications. Part I focuses on the introduction to machine learning. The author also attempts to promote a new design of thinking machines and development philosophy. Considering the growing complexity and serious difficulties of information processing in machine learning, in Part II of the book, the theoretical foundations of machine learning are considered, and they mainly include self-organizing maps (SOMs), clustering, artificial neural networks, nonlinear control, fuzzy system and knowledge-based system (KBS). Part III contains selected applications of various machine learning approaches, from flight delays, network intrusion, immune system, ship design to CT and RNA target prediction. The book will be of interest to industrial engineers and scientists as well as academics who wish to pursue machine learning. The book is intended for both graduate and postgraduate students in fields such as computer science, cybernetics, system sciences, engineering, statistics, and social sciences, and as a reference for software professionals and practitioners.

InTech Europe
University Campus STeP Ri
Slavka Krautzeka 83/A
51000 Rijeka, Croatia
Phone: +385 (51) 770 447
Fax: +385 (51) 686 166
www.intechopen.com

InTech China
Unit 405, Office Block, Hotel Equatorial Shanghai
No.65, Yan An Road (West), Shanghai, 200040, China
中国上海市延安西路65号上海国际贵都大饭店办公楼405单元
Phone: +86-21-62489820
Fax: +86-21-62489821

IntechOpen