Optimization and Design on Hotel Management Information System

Guo Chen¹, a

¹Xinyang College of Agriculture and Forestry, Xinyang, Henan, China 464000

a147426399@qq.com

Keywords: hotel management; system structure; resource optimization; k-means cluster; ID3 decision tree

Abstract. through the overall structure analysis and design of hotel management information system such as system software structure, system network structure and work procedure etc, this system realizes network management of hotel business, solves information communication and procession among departments. This paper analyzes characteristics and shortages of k-means cluster algorithm and ID3 decision tree algorithm, it also puts forward improved methods for its shortages and applies it into client analysis function module of hotel-client relation management system to identify the potential client group.

Introduction

As one of the industries opening to the outside world, hotel industry has stridden over the primary development stage and it is gradually going towards the mature stage[1, 2]. The gradual opening of market economy, quick development in tourism industry and promotion in living standard of citizens and transformation in consumption concept in China provide good conditions for quick development of hotel industry in China, gradual increase of residents in China and hotel consumption demand are the core driving force for industry development. With the quick development of information technology and computer hardware technology, the way of people dealing with information is slowly changing[3, 4]. Information technology is replacing the traditional information procession and information management tasks through its merits such as quick and correct information procession, friendly terminal interface to users etc, traditional hotel management appearance also has fundamental changes, manual management mode not only restricts management scope but also it is easily to occur mistake in the process of hotel management appearance, so it not only reduces service quality for customers, but also has great effect on hotel asset.

This paper makes observation, analysis and research on design and realization of information management system for one hotel, it presents the overall structure analysis and design of hotel management information system through system software structure, system network structure and system function component, it also establishes relevant work of intelligent self-service platform, makes charging and settlement according to usage condition. In this paper, it analyzes characteristics and shortages of k-means cluster algorithm and ID3 decision tree algorithm, as for its shortages, it puts forward the improved methods and applies it into client analysis function module of hotel-client relation management system to identify the potential client group.

System Design

The overall system structure design

The system design adopts modularized design and attributes complicated system into one module according to different functions. Every module has independent function, but all the modules make organization and form conforming entirety according to designated methods, so that it can complete the functions required by the whole system. Hotel management system mainly includes foreground operation, data management, and system management. Foreground operation mainly includes reservation registration, change registration, query, additional margin, clear, data management is mainly guest room management, adds guest room information and guest information, system
management includes user management and privilege management. The static design of system also should include background business, in this paper; I illustrate human resource management business and accounting management business, material supply management business. The package diagram of hotel management system is indicated by figure 1.

According to figure, we can see that system is mainly divided into clerk information management, menu information management, ordering information management, desk information management, supplier information management, consumption record management, client feedback information management etc. Foreground function module design includes guest room reservation management module design, guest staying management module design, guest room change management module design, guest check-out management module design; background function module design includes human resource management module design, accounting business module design, food service management module design, material supply module design and system management module design.

**System network structure**

System structure adopts distributed network structure of II grade, monitoring computer of every floor and guest room server adopts network topology structure based on C/S star shape, it uses TCP/IP to make communication, adopts quick Ethernet link and realizes it through Winsock. The lower machine is the level two of controller, it is responsible for site control and information collection, floor monitoring computer is upper machine, it timely monitors all the guest rooms in this floor, realizes correction of display and control parameters, information procession.
The E-R diagram among consumption entity is indicated by diagram 3, according to development of survives flow, relations among client, guest room and consumption are relatively simple. One guest room can be used by many clients, but one guest room only has one consumption list. It can implement payment by producing many orders according to order division and integration. One consumer list has one or many consumption lists.

Figure3 E-R diagram among each consumption entity
Each entity of goods management is doughnut-shaped operation in business. The purchasing order, purchasing order department and payment unit are one-to-many relations. One payment order can pay to many purchasing orders of one supplier, meanwhile one good can be provided by many suppliers, in-stockroom operation can make in-stockroom operation for many purchasing orders, meanwhile, the goods in one purchasing order can be put in storage for many times. The E-R diagram among goods entity is indicated by figure 4.

![Figure 4: E-R diagram among product entity](image)

**Work flow chart of hotel management information system**

The system work flow chart of hotel management information system mainly describes business flow chart of hotel management information system, DFD diagram of hotel management information system on the first floor. Business flow chart of hotel management information system mainly introduces relevant business and information procession made before clients check-in. When clients book hotel guest room online, it will produce reservation requisition form, check-in information form, change application form, check-out information form. Reservation requisition form, its main process is to confirm or cancel reservation, when clients cancel reservation, it will directly exit system, when clients confirm reservation, it will enter reservation information module, meanwhile it enters check-in information module, then summarizes client information (name, telephone, check-in room type etc) and reservation information. When clients need to change room, continued room and check out, it will respectively enter different modules. Change requisition form is mainly to summarize the usage information of guest room into the general module of information according to change room, continued room of clients. The check-out information form is mainly to deal with client requirement to check out, when guest room usage information of clients is summarized to the general module, the detailed flow chart is indicated by figure 5.
The DFD diagram of hotel management information system on the first floor is mainly a 2-level business process of client and administrator. When the client enters the system, they will first reserve a hotel guest room and submit a reservation requisition form. This requisition may have the possibility of rejection or acceptance. At this time, the administrator reviews the reservation requisition form submitted by the client. When the requisition form is accepted, the guest room will be allocated according to the client's demand. The client will check in and register check-in information, which will be stored in the database. When the client checks out, the settlement list will be entered, and the administrator will make settlements to generate receipt bills and return them to the client interface. During the entire process from client check-in to check-out, some guest rooms may be adjusted, at which time the system administrator should timely deal with the room adjustment requisition submitted by the client. If the review is approved, the administrator will notify the client of the type of adjusted room and arrange for the client to check in. The detailed business flow chart is indicated by figure 6.

Figure 5: Business flow chart of hotel management information system
Client resource optimization algorithm of hotel management information system

Data K-means algorithm of information resources

Choosing the proper data mining algorithm is the core for client acquisition. The algorithm choice should consider client data type and mining task. K-means is one of the basic division methods in cluster analysis, because it is reliable in theory with simple algorithm and quick convergence rate, it can effectively deal with large data set, so it is universally used. ID3 algorithm is the core algorithm in decision tree, it is simple and algorithm result is easily to be understood by client, so this paper adopts K-means algorithm to make detailed subdivision on client and provide one reference model for potential clients acquisition, it also adopts ID3 algorithm to establish reaction behavior model of potential client.

The aim of K-means algorithm is to divide data set into k clusters according to input parameter k. The algorithm adopts Iterative update method: in each turn, it respectively composes the surrounding points into k clusters according to k cluster center, while centroid of every cluster (that is the average value of all the points in the cluster, it is also geometrical center), which will be regarded as the cluster center of the next iteration. The iteration makes the selected centroid center much more close to the real cluster centroid, so the cluster effect will be much better. The cluster process is indicated by figure 7.

In every iteration, it allocates each data to the cluster center that closet to it, the time complication of this process is n(kd), here n means the number of the total data object, k means the designated cluster number, d means the dimension number of data object, after new classification is generated, it needs to calculate the new cluster center, the time complication of this process is n(d). So the total time needed by this algorithm in every iteration is O(nkd). This paper puts forward one kind of idea of using triangle with three unequal side law to reduce calculation time of every iteration by K-means algorithm for the shortages that K-means algorithm with larger expenditure in dealing with plenty of data.
Command $x_i \in X, \ d(c_k, c_j)$ as distance of 2 cluster center, 3 sides $d(c_k, c_j) \cdot d(x_i, c_k)$ and $d(x_i, c_j)$ compose one triangle, so it has the following:

$$d(c_k, c_j) \leq d(x_i, c_k) + d(x_i, c_j)$$  \quad (1)

If $d(c_k, c_j) \geq 2d(x_i, c_k)$, then it has the following: $d(x_i, c_k) \leq d(x_i, c_j)$, that is to say distance from $x_i$ to center $c_j$ is larger than distance of $c_k$. So under the precondition of $d(c_k, c_j) \geq 2d(x_i, c_k)$, it is unnecessary to calculate $d(x_i, c_j)$. The evolution of K-means algorithm is indicated by the following:

Step 1: give data set $X$ of $n$, command $I=1$, adopt $k$ primary cluster centers $c_j(I), j=1, 2, 3, ..., k$;

Step 2: calculate distance $d(c_j(I), c_j(I))$ between 2 cluster centers, of which $i=1, 2, ..., k, j=1, 2, ..., k$;

Step 3: suppose the current class of $x_i$ is $w_m$, calculate distance of $x_i$ and $w_m$ class center $d(x_i, c_m(I))$.

If $d(c_m(I), c_j(I)) \geq 2d(x_i, c_m(I))$ is false, then calculates $d(x_i, c_j(I))$;

If $d(x_i, c_j(I)) < d(x_i, c_m(I))$, then temporarily allocates $x_i$ to $w_j$, back to Step 2 to make circulation operation, finally classifies $x_i$ into cluster $w_m$. Of which $j=1, 2, ..., k; i=1, 2, ..., n; m=1, 2, ..., n$.

Step 4: $I=I+1$, calculate new function value $J_c$ of error square and norm function in the new cluster center according to formula (2)

$$c_j(I) = \frac{1}{n} \sum_{x \in w_j} x \quad (2)$$

$$J_c = \sum_{j=1}^{k} \sum_{k=1}^{n_j} \left| x^{(j)}_k - c_j(I) \right|^2 \quad (3)$$

Step 5: if $|J_c(I+1) - J_c(I)| < \xi$ is correct, then algorithm is ended, otherwise, command $I=I+1$, back to Step 2 to implement.

Make one comparison as for algorithm k-means 2 and k-means 3, recalculate cluster on the second circulation period, the time complication of these 2 algorithms are the same without distinction. But when it designates cluster in the first circulation period, the imported k-means algorithm obviously reduces calculation. Firstly consider the condition of sample point, in the k-means algorithm, the time
of calculating sample point to center point is \( k \), while in the k-means 3 algorithm, under the best
case, the time of calculating sample point to center point is 1, under the worst case, the time
of calculating sample point to center point is \( k \). Suppose the average calculation time of one sample
point in the first circulation period for every iteration. The total time complication needed by every
iteration in k-means 2 is \( O(nkd) \), the total time complication needed by every iteration in k-means 3 is
\( O(nkd) \), if there are too many sample points in the data set, which means when \( n \) is relatively larger,
the superiority of algorithm reveals.

**ID3 decision tree algorithm**

One biggest advantage of learning algorithm based on decision tree is that users are unnecessary to
learn about much more background knowledge in the learning process. Only training cases can use
way of nature-conclusion to present, it can use this algorithm to make learning. The key of decision
tree algorithm lies in how to choose one split nature to form the decision node of decision tree and
generate decision branch from the current node. When choosing certain split nature as decision node,
we usually hope that this nature can maximally reflect classification characteristics of training sample.
In the ID3 algorithm, choice if decision node nature applies the concept of entropy in information
theory as inspired function. In the choice method of this nature, it chooses nature with the maximum
gain nature as current division node. The selected node nature by this way can guarantee decision tree
has the minimum branch number, which makes minimum redundancy of decision tree. We will not
introduce the detailed calculation way in details. ID3 applies concept of information gain on choosing
split nature, the basic theory of algorithm is clear, which makes it simple of algorithm, so it is one
event example learning algorithm with practical value. However, because decision tree itself is not very
perfect, especially when it faces with many new problems in data mining, such as large-scale data ser,
use participation, knowledge intelligibility and usefulness etc, decision tree also faces with some
realistic problems, it needs further research to find better solutions.

In each selection process, it has the same value for nature \( A \), \( E(S_1, S_2, \ldots S_m) \), so information gain
\( \text{Gain}(A) \) is related to \( E(A) \), it is determined only by \( E(A) \). While the aim of \( \text{Gain}(A) \) is to
obtain the maximum classified information on training set \( x \) when it uses nature \( A \) as test nature, which makes
system entropy after division minimum(the smallest disorder degree). It only considers the
probability distribution condition on value and class of nature, this causes value of \( \text{Gain}(A) \)
maximally depend on the value of classified nature in training set. As for this problem, some people
puts forward concept of nature entropy, it uses probability distribution characteristics of nature value
to define nature entropy.

As for training set \( x \), suppose value of nature \( A \) is \( v_a = \{a_1, a_2, \ldots a_v \} \), \( x \) is the value of any sample in
nature \( A \) is \( a_i \), according to definition of information theory, the uncertainty of defining nature \( A \)
\( E_A \) is as follows:

\[
E_A(A) = \sum_{i=1}^{v} p(a_i) \log_2 (p(a_i)) \tag{4}
\]

Of which, \( p(a_i) \) is the probability of nature \( A \) taking \( a^* \), so it constructs one new inspired function:

\[
\text{AttIGain}(A) = E_A(A) - E(A) \tag{5}
\]

\( E_A(A) \) is only related to value probability and distribution of nature \( A \), while it has no relation to
current classification of training set. When the value probability and distribution of nature \( A \) are much
more average, the bigger value of \( E_A(A) \), the possibility of \( A \) being chosen as the best split nature is
much larger, on the contrary, the more concentrated distribution of \( A \) value probability, the smaller
value of \( E_A(A) \), the possibility of \( A \) being chosen as the best split nature is much smaller, on the
contrary, the more concentrated distribution of \( A \) value probability in all category, the smaller value
of \( E_A(A) \), the possibility of \( A \) being chosen as the best split nature is much larger. The advantage of
this is to organically combine distribution condition of objective probability of nature value with
probability distribution condition of value category, which avoids the condition that algorithm easily
to be caught in the local optimal.
The test demonstrates that the decision tree generated by inspired function is in the leaf node, height of tree all superior to the original ID3 algorithm. But the inspired function like this structure can not avoid deviating towards nature with bigger value. If we can introduce into the value number of nature and calculate the average gain of every value for every nature, in theory, it reduces the possibility of larger value being chosen, it corrects the value deviation problem of ID3 algorithm to certain extent. So inspired function is improved as follows:

$$\text{AttIGain}'(A) = \frac{\text{AttIGain}(A)}{N} \quad (5)$$

Of which, N is the value number with nature A. We take the maximum test nature of AttIGain'(A) as the best split nature.

Test simulation analysis on hotel information management system

The main development environment of hotel information management system adopted by this paper is as follows:

- **Hardware configuration**: Memory 32GB; Storage: DS4700 1TB; IBM X3850M2 XEON E7460*2 2.66GHZ;
- **Operation system**: Windows 2008 Server;
- **Data base**: ORACLE 10g relational data base;
- **Development tool**: Eclipse3.2;
- **Development structure**: Struts2.0 Spring2.0 Hibernate3.1;
- **Java compiler**: JDK1.5.0-12;
- **Web server**: Tomcat6.0

It uses every aspect of hotel managed by this system, guest room management, restaurant management and client information management belong to foreground management, and the content of background part has password management system, application system of cleaners and staff management. The staff management includes staff dispatch and modification of staff information etc. Management of client information includes modification of client information and query etc. Restaurant management includes restaurant of non-check-in client and restaurant of check-in client. Guest room management includes guest room reservation, cleaner application management includes which room needs cleaning, and setting interface of hotel guest room is indicated by figure 8.

![Figure 8: Setting interface of hotel room](image)

The management interface of restaurant is indicated by figure 9.
The test environment of hotel management information system is as follows: application server: Apache Tomcat 5.0, test and development tools: Java Build 2006: operation system platform: Windows Server 2008. The method on hotel management system adopts pressurization test, while the model on test adopts pressurization step by step. It starts 10 subsequent Vusers every 30 seconds, the total number of users is 200, it implements user login and logout operations, and it makes login request and operation one by one according to time interval set by users. The pressurization curve diagram is indicated by the following figure 10: the overtime time interval of http is set as 900 seconds. The required time of this test process is 44 minutes and 5 seconds. The biggest user number is 200. The throughput capacity from server end to client terminal is 1589536578 bytes, the click amount is 129606. There is no example in the system is failed, it is always successful. From here, we can see that capacity of hotel management system can achieve the predicted and realized design requirement.

Conclusion

Hotel management information system concerns whether one hotel can develop and expand, it concerns the future of hotel industry and it is the core for long-time development of hotel. At present, hotel industry is dedicated to developing much more effective management information system, which not only brings about updating in management for hotel, but also brings more economy benefit for hotel. Especially with the quick development in network information, most industries tend to develop E-business, it is the big trend for hotel industry, so this paper carries out design and realization on hotel management information system as for the trend of hotel industry at present, it
also analyzes characteristics and shortages of k-means algorithm and ID3 decision tree algorithm, it puts forward improvement methods and applies them into the client analysis function module of client relations management system in Junshan Hotel to identify the potential client group.

References