

Hospitality data mining myths

by Rob Law

Electronic database handling of business information has gradually gained its popularity in the hospitality industry. This article provides an overview on the fundamental concepts of a hotel database and investigates the feasibility of incorporating computer-assisted data mining techniques into hospitality database applications. The author also exposes some potential myths associated with data mining in hospitality database applications.

A database can be defined as consisting of “some persistent data that is used by the application system of a given enterprise.”¹ According to Date, persistent data are meant to be different in kind from other transient data, whereas the term “enterprise” refers to any organization. Fairlie, however, provides a more detailed and formal definition of a database. In Fairlie’s literature, a database is “a collection of data entries, which is organized on a computer by a software package, and which allows the user to relate, collate, summarize and reproduce such entries in accordance with any internally logical criteria.”² The natural advantage of database applications would be the personalization of products and services. This, in turn, maintains the direct contact with customers and establishes the loyalty through the creation of a two-side meritorious linkage.

A database also consists of multiple files for a specific organization. Each file contains multiple records for individual customers or products. A record (sometimes also known as a tuple) maintains multiple fields (or sometimes called attributes). These fields describe the behavior of a specific customer or product. Figure 1 demonstrates a sample hotel database which contains three files: Room File, Customer File, and Payment File. These files are used to show, for example, how the Island Shangri-La in Hong Kong might store its sales data.

Assuming Island Shangri-La stores the purchase records of accommodation services made by different guests, the files in Figure 1 can be utilized to draw on the database to decide special packages offering. The files representing these purchases would appear as shown in

Figure 1
A Sample Hotel Database

| Room File | | | | |
|------------------|------------------|-------------------|--|--|
| | Room Type | Name | | |
| | SK | Smoking king | | |
| | SQ | Smoking queen | | |
| | HS | Hospitality suite | | |
| | PS | Penthouse suite | | |
| | NK | Non-smoking king | | |
| | NQ | Non-smoking queen | | |

| Customer File | | | | |
|----------------------|----------------|------------------|--------------------------------|------------------|
| Name | Persons | Origin | Number of Nights Stayed | Room Type |
| Rob Jones | 1 | Permanent folios | 2 | HS |
| Kevin Adams | 2 | Weekend rate | 2 | NQ |
| Norm Smith | 1 | Corporate | 3 | NK |
| Rob Jones | 3 | Permanent folios | 3 | PS |

| Payment File | | | |
|---------------------|------------------|-----------------------|--------------------|
| Name | Room Type | Payment Method | Rate Reason |
| Rob Jones | HS | AE Gold Card | Package |
| Kevin Adams | NQ | Cash | Special rate |
| Norm Smith | NK | Visa Classic | Special rate |
| Rob Jones | PS | AE Gold Card | Package |

the appropriate fields of Figure 1. The database design in Figure 1, a relational database, is popular. The design is simple and well-defined, and can be implemented in most database management systems.

Suppose Island Shangri-La would like to offer a special package to visitors for three nights, in conjunction with Cathay Pacific Airway (the local airline company in Hong Kong), for a Penthouse Suite. Former Island Shangri-La customer Rob Jones, instead of Kevin Adams or Norm Smith (See Figure 1), would be the logical marketing target. This offer is made under the assumption that these people would be visiting Hong Kong when Island Shangri-La would be offering its special package. Based on this logic, Island Shangri-La would send a promotional brochure to Rob Jones's business/residential address (not included in Figure 1 due to the lack of space).

Data storage and representation in a relational database are not difficult to comprehend. Users can view, through various files, types of relationships and how these relationships can be handled. To demonstrate, one could find out the names of customers who had

stayed in a Penthouse Suite — Room Type PS — and their payment methods by searching through the Customer File and Payment File. To locate records, a user can use a query language such as SQL from IBM to join files together to perform a cross-file search. However, the information access and retrieval from a relational database is technical in nature, which might cause a usage problem for non-technical hotel managers.

Most hotel databases are designed in the relational format because of its long and stable history. However, researchers and hoteliers are also developing non-relational databases for marketing purposes in hotels.³

In general, the selective capability of a database makes it much more powerful than conventional mass marketing strategies. The primary purpose of applying databases is to set up a detailed profile of potential customers for marketing purposes. A goal of this approach is to allow a hotel, as well as other business firms, to aim at more appropriate customer groups. Using database marketing, target customers are well defined and smaller in scale to meet certain criteria. Some examples of selection criteria include family background, career, and purchasing history. The database marketing format allows decision makers to target specific customers for advertising campaigns by issuing Database SQL commands to ask, for instance, "Who has lived in a Penthouse Suite in Island Shangri-La and paid by AE Gold Card?" Database marketing provides a better aim, selectivity, personalization, and customer response rate at a lower cost compared to conventional mass marketing strategies.

Hospitality data mining is a necessity today

The process of gathering and comprehension of useful information from a set of raw data is a necessity for every business to achieve competitive advantages in the present world. Naturally, this situation applies to the hotel industry as well. Similar to the scientific and engineering disciplines, an intrinsic problem faced by hotel managers in today's business environment is the large amount of raw data carried. Hoteliers, usually non-technical professionals, may have difficulty comprehending the increased complexity of raw data. In general, managers often find retrieval of useful information from raw data to be a difficult, slow, and mysterious process. Traditional manual information retrieval methods include desk reviewing of raw data, dynamic tracing with keyword recognition, and *ad hoc* searching for various patterns. However, managers are often flooded with raw data, which are sometimes imprecise, incomplete, uncertain, redundant, and contradictory. This problem gets worse as data quantity increases. In other words, the oversupply of raw data often distracts the managers from locating the useful information during managerial decision making.

At present, various database management software packages are available for information storage and retrieval. Some examples of these database management tools are the DBASE products, Ingres, and Oracle. In spite of the previous successful applications in information storage and retrieval, database information retrieval (sometimes called database knowledge discovery) is limited in its functions and capability. To illustrate, a guest history database is simply a static collection of data with no indication of relationships between guests and variables relating to them individually. More importantly, a database is a solved problem and is therefore not interesting to generate new, surprising, and interesting concepts. Routine decisions are suitable for database information retrieval but not the dynamic hotel business decisions.

Data mining is becoming increasingly popular

In view of the comprehension problem of excessive raw data in the business environment, researchers in computer science developed an approach named data mining to retrieve practical, previously unknown, and implicitly useful information from databases. Technical approaches borrowed from computer science, in particular artificial intelligence, are incorporated into the data mining process. The common artificial intelligence methodologies which are utilized in data mining include large parallel array processors, fuzzy logic, and neural networks.⁴ In fact, most models now used in data mining tools are not fundamentally different from the traditional quantitative model-building techniques. Instead, data mining methods are natural extensions and generalizations of the analytical techniques which have existed for decades. Neural networks, for example, are simply a special case of regression.⁵ K-nearest neighbor, another data mining method, has been used for more than 30 years. All these data mining models, similar to the century-old regression techniques, find the relationships between a set of independent (or profile) variables and a set of dependent (or outcome) variables. What is new in data mining is that people are now taking these techniques from academia and research centers and applying them to more general business problems. The increasing availability of data and inexpensive computer storage and processing power also act as the catalyst for the increasing popularity of data mining.

The central idea behind data mining is to identify the hidden relationships and patterns which are beyond visual capture by human eyes. Data mining would be useful for an organization to analyze and exploit opportunities, especially in mature or declining markets. The theoretical background of these techniques can be found in a number of sources.⁶ Useful information retrieved from data mining software will be stored in a form which represents the discovered pattern or

knowledge. The mined information or knowledge will then be used by end users (usually the non-technical business managers) to assist decision making.

Various earlier work has been performed to study the relationship between artificial intelligence and hospitality management.⁷ However, to date, not much previous work has been attempted to incorporate data mining into hotel databases.

In the hotel context, data mining software can be used by a house-keeping manager to identify the relationship between the specified floors and the inspectors and their attendants serving these floors. Similarly, a F&B manager, utilizing a data mining tool, can analyze the relationship between sales volume, number of covers, inventory control methods, and number of function rooms. A front desk manager can use a data mining system to detect the associations among the number of registration windows, employees per shift, average/maximum number of check-outs per day, and average/maximum number of registrants per day.

In short, data mining can be applied virtually everywhere in a hospitality environment. Processes from restaurants, rooms division, and concierge can be understood, studied, and improved. Data mining tools can be applied to such diverse areas as human resources and accounting departments. Data mining is simply a technique to detect or discover any hidden relationships or patterns in a hotel database to help hoteliers make better business decisions.

Myths do exist in hospitality database applications

There exist many myths associated with hospitality data mining. The most common ones are the following:

Myth 1: The more data that you supply for the data mining tools, the more effective the data mining technique will be. Therefore, all existing data from every hotel database, regardless of the relevance, should be brought into the data mining process.

While data mining tools can generate more patterns with more data input, patterns, however, are useful only if they contribute to the topic at hand. Otherwise, these patterns could be worse than useless as they may distract the hotel managers from locating the pertinent information. A database may contain a lot of information about a customer but nothing about other customers who are actually related. For instance, a hotel may have information about how a guest uses some specific credit cards, but not about how these credit cards are used by other guests.

Worse still, adding data items with no/little information content to the data mining process will actually reduce the analyzing power of the database. The inclusion of an irrelevant data item or putting multiple measurements of the same item can lower the usefulness of

data-mining results. A likely example would be if birth date and age of a guest are both included in the analysis, the computer results may return an equal weight to both variables. This, in turn, lowers the forecasting power of both variables.

Myth 2: Only huge databases are worth mining. Patterns or useful information cannot simply be observed in small databases.

It is correct that many business data mining applications involve large databases and many data mining techniques and tools are thus built to deal with large databases. Still, a small or medium-sized database can generate useful knowledge which may be beyond visual catching. For example, the hotel occupancy rate may depend strongly on the day of the week or time of the year. A modest-sized guest history database consisting of day, time, and sales would show this valuable information, giving the hotel manager some ideas of the relationship magnitude, and facilitate the planning of marketing and staffing.

Hoteliers can collect different data altogether or collect data differently. It is strongly recommended that before a massive database is put into use, it should be tried out with some simple analyses while it is still moderate in size.

Myth 3: Once the database is available, without being instructed what to do, the data-mining software automatically finds the needed patterns.

Within a given application domain, data mining is effective to retrieve valuable information. Although data mining software can investigate the database and detect patterns or relationships, a user is still needed to direct the software toward a specific goal. Simply giving the data mining tool a mailing list from a guest history database and expecting the tool to improve the response rate of a direct mailing campaign will not be an effective approach. Hoteliers need to be more specific in their goals. To demonstrate, to increase the financial value of the mailing list responses, hoteliers can emphasize guests who have previously stayed in suites for more than three days. Similarly, to increase the response rate of direct mail, hoteliers should emphasize guests who have responded to previous mailings.

Myth 4: Since information from unused data is not included, the model built on a sample of a database using a data mining approach might be incomplete and ineffective.

Statistics theory asserts sampling is to maximize the amount of knowledge gained for every unit of effort being used. Sometimes it is necessary to sample since not all data are relevant to the problem at hand or reflect the modeling population. A database with historical data that reflect conditions, such as the guest dining information during the Vietnam War, which may no longer be relevant, will render inappropriate model building for future hotel marketing direction.

Additionally, building a full-scale database for data mining is impractical. For example, a F&B manager wants to learn about

customers' satisfaction with a new restaurant, but it takes an hour to manage a guest survey. The F&B manager would rather limit the data collection to a sample of customers.

Myth 5: Since data mining involves some complicated technical processes, data mining software would be very difficult to use.

For most non-technical hoteliers, the technicality associated with a data mining model could be very complicated and difficult to understand. However, many market-available data mining systems have made the application process fairly easy. Very often, outputs from data mining tools are in a form of simple tables and graphs. However, these tables and graphs can mean a lot to the hotel business. If there are difficulties in using data mining tools, most of these difficulties are from data organization and data preparation issues. For instance, users always come across the difficulty of deciding which variables to include and how to arrange the variables.

Myth 6: Data mining tools can produce surprising results which will completely change the hotel business.

Useful patterns generated by data mining software can contribute to the improvement of new and well-established hotels in knowledge gained about hotel guests. This knowledge can always cause positive evolutionary changes, instead of revolutionary changes, to the hotel business. Small changes, while compounded over a certain period of time, can lead to a substantial increase in business. However, data mining in the hotel context can occasionally discover some "interesting" facts, such as the fact that Oriental guests are always the "big spenders" in hotel casinos. Data mining systems are therefore good for forward-looking hoteliers.

Myth 7: Data mining is just another fad which will shortly fade.

Standard business practice is still in operation with the existence of data mining tools. Hoteliers should remember that data mining systems are developed to assist, among others, hotel managers, but not to replace anybody. In addition, the name may change in the future, but data mining as an important application area in business does not end. Data mining is simply another advance in the current research progress.

Expectations about data mining techniques abound

In the hospitality industry, and probably other service industries, a typical technological cycle always goes through five stages:

- excitement or enthusiasm for a specific claim
- misunderstanding or ignorance of the technology's potential benefits and actual capabilities
- adoption by hoteliers without enough background preparation and training

- frustration, sadness, disappointment, complaints, and divorce from the technology
- a return to business as usual.

To deal with a data mining project with whichever data mining software, hoteliers should avoid unrealistic expectations and disappointments. To ensure the success of a hotel data mining project, hoteliers have to understand the facts and not have any false hopes. As long as data can be gathered, data mining would be useful. However, a cost/benefit analysis may show that the time and effort being spent on data mining actually exceeds the return.

For example, a hotelier anticipates by administering one more piece of information about hotel guests, say, annual income, that room occupancy rate will increase 20 percent. However, the hoteliers also know that by sending mail to twice as many people, the room occupancy rate will increase 20 percent. If gathering, storing, and retrieving the extra data electronically is more expensive than sending the mail, it would be more realistic for hoteliers to send the extra mail rather than handle and mine the extra data.

References

- ¹ C. J. Date, *An Introduction to Database Systems* (New York, N.Y.: Addison-Wesley, 1995), 9.
- ² R. Fairlie, *Database Marketing and Direct Mail - A Practical Guide to the Techniques and Applications* (Herts: Kogan Page, 1995), 53.
- ³ R. Law, "Integrating Database Marketing and Decision Support Systems: A Study of the Hong Kong Hotel and Tourism Industry," *International Journal of Management* (1998), in press.
- ⁴ G. Lennox, "An Intelligent Approach to Integrating Marketing Databases with Other Technologies," *Emerging Information Technologies - A Marketing Opportunity* (1993): 24-29.
- ⁵ W.J. Frawley, G. Piatetsky-Shapiro, and C.J. Matheus, "Knowledge Discovery in Databases: An Overview," *AI Magazine* (Fall 1992): 57-70.
- ⁶ S. Sestito and T. S. Dillon, *Automated Knowledge Acquisition* (New York, N.Y.: Prentice Hall, 1994).
- ⁷ E. Crichton, and D. Edgar, "Managing complexity for competitive advantage: an IT perspective," *International Journal of Contemporary Hospitality Management* 7, no. 2/3 (1995): 12-18; R.C.H. Law and F. Go, "The Application of Database Marketing to Achieve Competitiveness in the Information Age," *Proceedings of the Academy of International Business Southeast Asia Regional Conference, Dunedin and Queenstown, New Zealand* (June 1996): 435-438.

Rob Law is an assistant professor in the Department of Hotel and Tourism Management, The Hong Kong Polytechnic University, Hung Hom, Kowloon, Hong Kong. The author acknowledges his appreciation for the financial support provided by the Hong Kong Polytechnic University (under contract #: 351/589).