ICEB 2002 Proceedings

International Conference on Electronic Business (ICEB)

Winter 12-10-2002

Development of a DSS to Estimate the Sales for the Retailing Industry in Taiwan

Ching-Chin Chern

David Ming-Huang Chiang

Ting-Wei Ho

Development of a DSS to Estimate the Sales for the Retailing Industry in Taiwan

Ching-Chin Chern
Dept. of Information Management
National Taiwan University
50, Lane 144, Sec. 4, Keelung Road
Taipei 106, Taiwan
Chern@chern.im.ntu.edu.tw

David Ming-Huang Chiang
Dept. of Business Administration
National Taiwan University
50, Lane 144, Sec. 4, Keelung Road
Taipei 106, Taiwan

Ting-Wei Ho
Information System Dept.
Computer & Communications
Research Laboratories, ITRI
Bldg 14, 195-11 Sec. 4, Chung Hsing Rd.
Chutung, Hsinchu 310, Taiwan

Abstract

An algorithm is constructed in this study to estimate the market sizes of daily commodity in Taiwan based on the sampled sales information provided by retailer chains. Though retailer chains provide sampled sales information from only small portions of their retailing stores, they expect to receive more valuable processed information from that. As result of this research, a DSS is proposed to compute value-added information from this joint sales information database, namely the estimation information. Through certain public accessible data such number of stores by each chain and retailers' financial reports, the sampled sales information can be transferred to the market size information of each item in Taiwan. Two similar algorithms are constructed for convenient stores and supermarkets/hypermarkets separately. A simple integration method is used to combine these results. Finally, a DSS is built based on these estimation algorithms and is implemented successfully.

1. Introduction

Sales information is always the important fundamental information that the decisions of the future sales and inventory policies were based on. However, not all the retailers have the ability to utilize information technology. Some retailers are unable to process and analyze the sales information further due to the lack of investment in capacity, capital, and people. Therefore, Department of Commerce in Ministry of Economic Affairs in Taiwan support a project initialized by Computer and Communications Research Laboratories (CCL) of Industrial Technology Research Institute (ITRI) to analyze and integrate the demands from all the retailers regarding to the sales information for the retailing industry in Taiwan. After analyzing the demands, CCL/ITRI setup Distribution Information Service Center, or DISC, to study and fulfill information requirement of retailers in Taiwan in May 2000. DISC collected sampled sales information from the retailers and constructed a joint sales information database, called "Retailing Sales Information Sharing System" or RSISS. There are currently more than 100,000 retailing items with Universal Product Code (UPC) in this database and all of them have been classified with three levels of classification codes.

As more and more retailers agree to cooperate in this project, the joint sales information database of RSISS is growing with more and more data. The retailers who joint this project is expecting to receive more value-added information such as estimation, prediction, etc in the future. In light of this expectation, DISC endows a project to construct a DSS that estimates the sales of retailing industry in Taiwan based on the sample sales data. This DSS is constructed and implemented successfully through the cooperation and efforts of DISC, Department of Information Management and Department of Business Administration in National Taiwan University.

2. Literature Review

As mentioned in Andreasen [1] and Rapp etc [12], the research in Marketing in the past usually concentrated on certain subject and solved specific problem from scenario analysis, to problem formulation, to data collection, to result finding. However, this type of research ignores the supply-demand environment issues and the relationship of marketing activities among different chain stores. The study in Marketing Research Systems (MRS) [4] or Marketing Information Systems (MKIS) [2] arises for this reason. Through collecting customer and sales information, a sales and marketing related database can be constructed to provide continuous and relevant sales/marketing reports. Morris etc [11] surveyed the views of 101 marketing managers toward MRS and found that they think positive of MRS though these managers don't exactly know what computer technology would affect them. Li [8] tracks the development of MKIS among the Fortune 500 companies in 10 years. He concludes that the usage of MKIS is getting more and more complicated and matured but the marketing managers are still unsatisfied with the reports and information provided by MKIS. Talvinen etc [15] also survey the senior marketing managers of 156 wholesalers in Finland and find that MKIS plays a crucial role in marketing policy making.

Constructing a complete MRS or MKIS needs consistent and long-term efforts as well as computerized capability to clean, integrate, classify, and transform data. Talvinen [14] suggests a specific system architecture of a MKIS with important processing subsystems such as decision support, planning, and internal/external control. Amaravadi etc [2] think the future trend of MKIS is Intelligent Knowledge System with marketing/sales knowledge base and decision support function. With a complete sales/marketing database, more and more researchers focus their efforts on the decision support functions that could be added on the system. Higby etc [6]

The Second International Conference on Electronic Business
Taipei, Taiwan, December 10-13, 2002

found that information after cleaning, classification, and computation is much more valuable to marketing managers in terms of decision making. Li etc [7] analyzes the utilization of MKIS in Fortune 500 companies and finds that marketing managers commonly use MKIS/Decision Support System/Expert System in their decision making and are more interested in value-added information such as sales forecast and market share estimation. Li [9] surveys the conditions of using MKIS in 1000 small US companies and finds a relationship between sales and probability of MKIS utilization. In other words, the larger the sales of a company, the more possible the company has a MKIS. As to how the company applies MKIS to marketing policies, 67.7% are for pricing strategy, 65.8% for new product evaluation, 61.8% for advertisement media selection, 52.6% for product out-of-market, 46.1% for budget, and 43.4% for salesperson allocation. Li etc [10] found that the support of MKIS to management is in 4 aspects: planning (44%), control (19%), organization (15%), direction (14%), and personnel (8%). They also think that more and more IT specialists joint the sales/marketing department to help salesperson or marketing personnel obtaining more value-added information without relying too much on MIS/IT department.

Adopting mathematical modeling and statistics method is also a future trend of MKIS to provide more valuable decision support functions. Among all the value-added information, sales estimation and forecast is the most important and common one provided by MKIS [7, 16, 17]. Marketing managers from all types of industries are all eager to know the sales estimation of a single product or the whole company, or the estimation of future market shares. Higby etc [6] also find that 66% of MKISs provide support on sales estimation/forecasting. The conclusion can also be found in Li etc [7]. Current research on data mining also showed the same result. Berry etc [3] think that one of the important functions provided by data mining is estimation followed by prediction. Concluded from the above reviews, estimation is a very important value-added information provided by MRS or MKIS. Therefore, this research proposes to build a DSS upon a joint MKIS, namely RSISS, to estimate the retailing market size of each individual item in Taiwan.

3. Problem Description

The problem faced by DISC is to estimate the market size of every individual product with UPC based on the sampling sales information provided by the retailer chain stores. There are three major types of retailing stores in Taiwan: convenient stores, supermarkets, and hypermarkets. Old-fashioned grocery stores now exist only in the remote areas and take up only a very small part of the total sales of the retailing industry in Taiwan. They are usually not equipped with IT capability and thus, are not able to provide digital sales information to DISC. Six major chains of convenient stores capture about 95% of the convenient store market in Taiwan and joint the project of Retailing Sales Information Sharing System (RSISS) founded by DISC in May 2000. As to the supermarket chains, major players capture more than 50% of the supermarket sales in Taiwan and joint the project of RSISS a year later than the convenient-store chains in May 2001. More supermarkets under negotiation are interested in joining the project. It is estimated that 85% of the supermarkets will be included in this project in the end of 2002. The negotiation of hypermarkets' cooperation in the project is currently under way. A full-scale cooperation from the major hypermarkets is expected to begin in July 2003.

With so many partners in the project, it is impossible to process all the sales information from all the stores. As to the end of May 2001, there are more than 6,250 convenient stores, 570 supermarkets, and 100 hypermarkets in Taiwan. On the average, each convenient store carries more than 4000 items with UPC while each supermarket carries more than ten thousand UPC items and each hypermarket carries forty thousand UPC items. Not only is it impossible to process such large scale of sales information, it is also unrealistic to ask retailer chains to submit all the sales information of all stores. To balance the information-sharing scale, each chain of convenient stores samples sales information of 100 stores. As to supermarkets, most of the supermarket chains have less than 40 stores except Wellcome. Therefore, Wellcome provides sales information of 40 stores out of its 105 stores while others provide sales information of all their stores. From the geographic point of view, the scattering of the stores around Taiwan is not balanced. DISC has divided Taiwan into three geographic areas: North, Center, and South. More than 60% of the retailing stores are located in Northern Taiwan. As result of negotiation, each convenient store chain provides sales information of 40 stores from north, 20 from center, and 20 from south if there are enough stores in each area. For supermarkets, Wellcome provides sales information of 20 stores from north, 10 from center, and 10 from south while others provide sales information of all their stores based on the locations. Random sampling is not applicable here since the scales of the convenient store chains vary significantly. To balance the differences among sampling stores, DISC requires the samples to be taken from the top 50% list for all the convenient store chains and Wellcome supermarkets. From the above sampling process, the problem of estimating sales faced by DISC can be depicted as Figure 1. In other words, the sales estimation of each individual item with UPC in Taiwan is computed based on the sampled sales data collected each month by DISC. The sales estimation can be seen as a summarized figure or be analyzed from different points of views.

4. Estimation Algorithm

Two separate but similar heuristic algorithms are constructed to estimate the market size of convenient stores and supermarkets/hypermarkets. A simple computation module is then used to combine these two results and provides the final estimation of the entire retailing market as seen in Figure 2. These tow heuristic algorithms are very similar in terms of estimating the market sizes but different in the analysis dimensions. Two types of analysis dimensions are applied: item and store characteristics. From item point of view, some items are seasonal or regional while others are not. In order to aggregate the similar effect, items are grouped into three levels of categories: first-layer, second-layer, and third-layer. Through the years, DISC has developed a standard classification mechanism to the items in RSISS. In this study, the same classification method is adopted. For the store characteristics, geographic areas, living zones of convenient stores, and sizes of supermarkets and hypermarkets are all the related analysis dimensions.

Six steps are involved in the algorithm of estimating the market size of convenient stores for each item as seen in Figure 3 and are elaborated below.

- **Step 1**: Compute the average sales at each store of each chain: The reason to compute this average is that some chains might not be able to provide sales information for the stores in all areas. However, those sales will have to be included in the estimation. Therefore, other chain K will be used as the reference to help estimating sales in those areas without any sales information. However, the scale of reference chain is different from the chain to be estimated in terms of sales. To adjust the difference in scales of different chains, average sales at each store of each chain will have to be calculated first.
- **Step 2**: Compute the average sampling sales of each first-layer category of each chain in each area: The reason to compute this average is due to the customer’s behavior differences among different areas. However, to compute the average for each individual item is an impossible task due to the enormous amount of item information. In order to reduce the amount of calculation efforts, first-layer category is used instead of individual item. To obtain the average, compute the total sales of sampling stores for each first-layer category of the reference chain of Chain R in reference area of area A first. Then compute the ratio of average store sales of Chain R in area A over the average store sales of its reference chain and area and adjusting the ratio by number of stores sampled by chain R and its reference chain.
- **Step 3**: Compute the percentage of the total sales for each first-layer category of each chain occurs in each area after normalization: The percentage of total sales at chain R occurred in area A of chain R can be collected from retailers. However, it is for the summary of all items instead of individual items. However, to calculate the percentages for individual items is a difficult job due to the same reason mentioned above. Since products in the same first-level category possess similar characteristics, this percentage is computed for each first-level category. To get the percentage, compute the total sales of each first-layer category at sampling stores of each chain in each area. Sum up the total sales of every first-layer category for each chain next. Then, calculate the percentage by dividing the above two figures. However, the sum of all percentages for each first-layer category in each area of each chain might not be 1. Thus, the percentages need to be normalized to make the sum of percentages equal to 1.
- **Step 4**: Compute the adjustment factor of the total sales for each first-layer category of each chain occurs in each zone after normalization: The purpose of this step is similar to step 3 but from different angle, namely living zones. To attain this adjustment factor, compute the total sales of sampling stores for each first-layer category of each chain in each living zone first. Sum up the total sales of the first-layer category J for chain R next. Then, compute the percentage for first-layer category J of chain R adjusted by the scattering of stores.
- **Step 5**: Compute the estimated total sales amount of each item in each living zone and each area: Compute the total sampling sales of each item in every zone Z of every chain first. Then, divide it by the percentage of the total sales at chain R occurred in the sampling stores and multiply the result by the adjustment factor computed in step 4 to obtain the total sales of each item in each zone of each chain. Multiply the percentage obtained in step 3 to the above result to get the sales of each item in each zone and each area of each chain. Finally, sum up the total from each chain to be the total sales amount of each item in every zone and every area.
- **Step 6**: Compute the estimated total sales amount of each first-, second-, and third-layer category in every zone and every area: To do this, simply add up the result from step 5 for all items belong to each first-, second-, and third-layer category in every zone and every area.

Certain assumptions have to be made in order to finish this algorithm and are elaborated as follows:

1. The sales pattern of each item in the same first-layer category is similar. Under this assumption, each item can apply the same percentage of the total sales for each first-layer category of each chain occurs in each area after normalization.
2. The sales pattern of each store in the same area of the same chain is similar, which implies that the

average sales amount of each store in the same area of the same chain is also very similar.

3. The sales pattern of each store in the same living zone of the same chain is similar, which implies that the average sales amount of each store in the same living area of the same chain is similar, too.

Six steps are also involved in the algorithm of estimating the market size of supermarkets or hypermarkets for each item as seen in Figure 4. As comparing with the algorithm for convenient stores, the differences lie in the analysis dimensions. In the algorithm for convenient stores, living zone is a very important analysis dimension, while size of store is used in the algorithm for supermarkets. Two different sizes are adopted in the algorithm: large and small. Stores with spaces larger than 661 square meters (7200 square feet or 200 pings) are defined as large while else are small.

After the previous two algorithms are run, the results are imported and combined by the following steps:

- **Step 1**: Compute the total sales amount of each item in each area
- **Step 2**: Sum up the total sales of each item in each area by $\Sigma_i \beta_i$*(sales estimated for retailer type i for each area) where β_i is the percentage of the sales of retailer type I included in RSISS.
- **Step 3:** Compute the estimated total sales of each first-, second-, and third-layer category in each area.

In the above algorithm, several parameters are assumed known and constant. However, difficulties arise when started to collect these needed parameters. The total number of stores of each chain in each area is public data and can be collected from government reports each month. The percentage of the total sales at chain R occurring in the sampling stores of chain R or P(R) is the most important ratio that needs to be gathered but all the retailing chains refuse to provide this information because of ferocious market competition. The percentage of total sales at chain R happened in area A of chain R or W(R, A) should be provided by all the retailing chains but is also rejected for the same reason. Therefore, an evaluation process is proposed in this section to compute the two important ratios, P(R) and W(R, A).

- Evaluation of P(R): By definition, P(R) is the percentage of the sales at the sampled stores in the total sales at all stores of chain R each month. Sales information at the sampled stores of chain R is collected every month. Therefore, the total sales amount at the sampled stores of chain R can be easily computed every month. The total revenue of chain R can also be obtained from the companies' monthly financial reports. However, several items such as prepared food, fee-collecting services, etc in revenues should not be included because they are not sales of UPC items. About 15 to 20 percent of the revenue should be excluded from the total sales in the financial reports used to compute P(R). For each chain, P(R) is equal to the total sales amount at the sampled stores divided by 85% or 80% of the total revenue found in the financial reports.
- Evaluation of W(R, A): To estimate W(R, A), which is the percentage of total sales at chain R happened in area A of chain R, numbers of stores in each area for each chain is collected every month. By interviewing the chain managers, it is known that on average, the sales of a store located in the south or central area is about q% less than the sales of a store located in the north where q is usually between 10 to 15. Therefore, W(R, A) can be computed on this base.

5. System Structure and Implementation

The position of this Estimating DSS in the RSISS is shown as the gray-fill box in Figure 5. The system is built on MS Server 2000 environment with MS SQL as database server and JDE as the development tool. The architecture of servers used in this DSS is shown in Figure 6. The system is built and tested on a PC server with Pentium IV 1.7GB CPU and 1GB memory. The testing data is from convenient stores and contains more than 2.4 million records. The computer computation time each month is less than 30 minutes because of efficient algorithm but the storage space requirement is enormous. The estimation of 2-year sales history occupies 40 GB of hard disk space. The estimating DSS is finished and implemented in December 2001. It started the estimation process since March 2002. Totally, 24 months of history sales data of convenience stores has been imported to the system. The estimation has been made for each item each month. Drilling down and summing up OLAP functions are also added to the DSS so users can query the estimation from item, classification, area, or living-zone point of views. The DSS is also built with web enabling capability so users can easily access the results through internet browsers as seen in Figure 7.

Although verification procedure has been carried out carefully to make sure the execution of the algorithm accurately, no public data exists to validate the result of estimation. The estimation result has been shown to the participants of the project including the convenient chain stores, supermarket chains, and hypermarket chains. In June 2002, a meeting is held among DISC, retailers and the manufacturers to show some estimation results from this system. The purpose of this meeting is also to validate the estimation results through manufacturers. The manufacturers all showed great interests in obtaining this estimation result mainly because there is no other source to provide such census retailing information. It is the first time that the manufacturers have the opportunity to be able to gain some insight of the local market for each individual product. To show some impact, the sales of a famous tea drink (200 ml) reaches 200 millions of NT dollars (more than 6 millions US$) alone last year in convenient stores market. To everyone's surprise, its biggest group of consumers is student since it showed the biggest sales in school zone. The estimation also shows that the market of tea drink is fierce competed among several large local manufacturers. With such valuable information, the manufacturers are willing to pay for this information but the scales and levels of services are still under negotiation.

The DSS for supermarkets and hypermarkets are also finished and implemented at the same time. However, the import of the sales information from these two types of retailers has not been completed until July 2002. The system is under testing and is expected to be on line at the end of this year. Because of the sensitivity and confidentiality of the sales data, it is prohibited to show the user interfaces and results of this estimation DSS here.

6. Conclusion

This study proposed an algorithm to estimate the market size of product with UPC in Taiwan based on the sampled sales information provided by retailer chains. By sampled sales information, it means that each retailer chain provides sales information from only a small portion of its retailing stores. Through certain public accessible data such number of stores by each chain etc, the sampled sales information can be transferred to the market size information of each item in Taiwan. A DSS is built based on this estimation algorithm and is implemented successfully. In the future, the estimation information can be further processed to provide more value-added information. One of the extensions will be to adopt some forecast models such as time series analysis and economic/environmental variables. From current estimation results, it is clear that zones, areas, and classifications of items all have impacts on sales. For example, sales of tea drink in school zone show very significant seasonal effect while sales of coffee on the other hand demonstrate no seasonal effect at all. In the future, different prediction models can be used for items in different classifications, zones, and areas.

Reference

[1] Andreasen, A. R., "Backward Market Research," Harvard Business Review, 1985, pp. 176—182.

[2] Amaravadi, C. S., S. Samaddar, and S. Dutta, "Intelligent Marketing Information Systems: Computerized Intelligence for Marketing Decisions Making," Marketing Intelligence & Planning, Vol. 13, No. 2, 1995, pp. 4—13.

[3] Berry, M. J. A. and G. S. Linoff, "Data Mining Techniques: For Marketing, Sales, and Customer Support," Wiley, 1997.

[4] Curry, D. J., "Single-Source Systems: Retail Management Present and Future," Journal of Retailing, 1989, pp. 1—20.

[5] Farnum, N.R. and L.W. Staton, "Quantitative Forecasting Methods," PWS-Kent, Boston, MA, 1989.

[6] Higby, M.A. and B. N. Farah, "The Status of Marketing Information Systems, Decision Support Systems and Expert Systems in the Marketing Function of U.S. Firms," Information & Management, 1991, Vol. 20, pp. 29—35.

[7] Li, E. Y., R. McLeod, Jr., and J. C. Rogers, "Marketing Information Systems in the Fortune 500 Companies: Past, Present, and Future," Journal of Management Information Systems, Summer 1993, Vol. 10, No. 1, pp. 165—192.

[8] Li, E. Y., R., "Marketing Information Systems in the Top U.S. Companies: A Longitudinal Analysis," Information & Management, Vol. 28, 1995, pp. 13—31.

[9] Li, E. Y., R., "Marketing Information Systems in Small Companies", Information Resources Management Journal, vol. 10, No. 1, 1997, pp.27-35.

[10] Li, E. Y., R. McLeod, Jr., and J. C. Rogers, "Marketing information systems in Fortune 500 companies," Information & Management, September, Vol. 38, 2001, pp.307-322.

[11] Morris, M. H., A. C. Burns, and R. A. Avila, "Computer Awareness and Usage by Industrial Marketers," Industrial Marketing Management, 1989, Vol. 18, pp. 223—232.

[12] Rapp, S. and T. L. Collins, "The New Direction in Advertising and Marketing Strategy," McGraw-Hill, New York, 1987.

[13] Schwedelson, R., "New Wave Database," Direct Marketing, 1998.

[14] Talvinen, J. M., "Information Systems in Marketing: Identifying Opportunities for New Applications," European Journal of Marketing, Vol. 29, No. 1, 1995, pp. 8—26.

[15] Talvinen, J. M. and T. Saarinen, "MkIS Support for the Marketing Management Process: Perceived Improvements for Marketing Management," Marketing Intelligence & Planning, 1995, Vol. 13, No. 1, pp. 18—27.

[16] Taylor, B. W., "Introduction to Management Science," Englewood Cliffs, N.J.: Prentice Hall, 1996.

[17] Wilson, J. H. and B. Keating, "Business Forecasting," Boston, MA: Irwin, 1990.

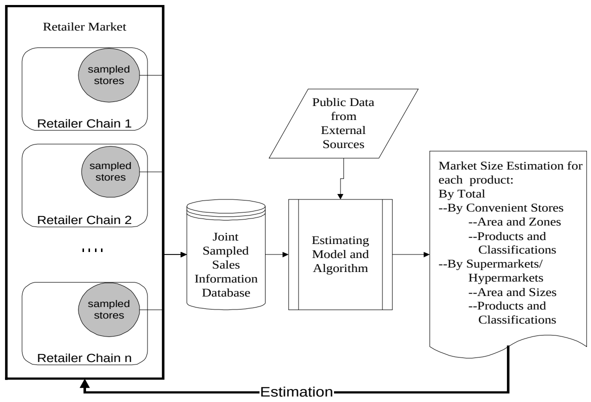

Figure 1: Problem of Estimating Sales

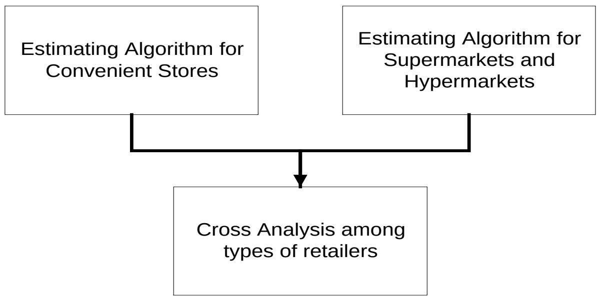

Figure 2: The Estimation Process

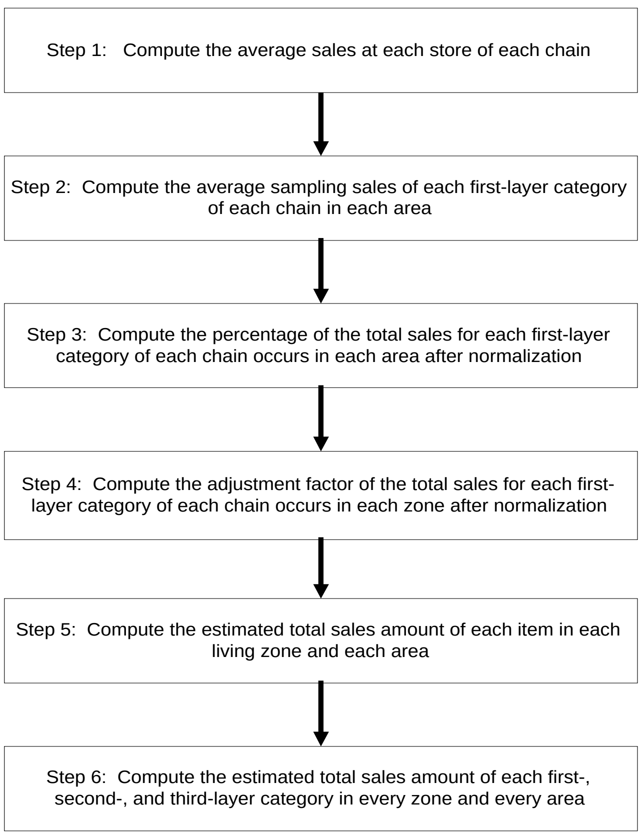

Figure 3: Algorithm for Convenient Stores

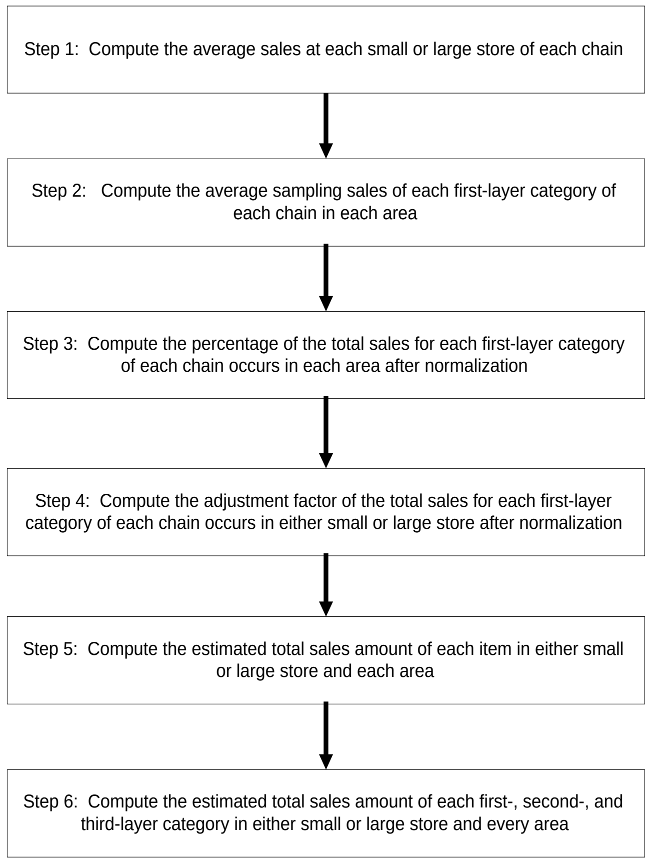

Figure 4: Algorithm for Supermarkets or Hypermarkets

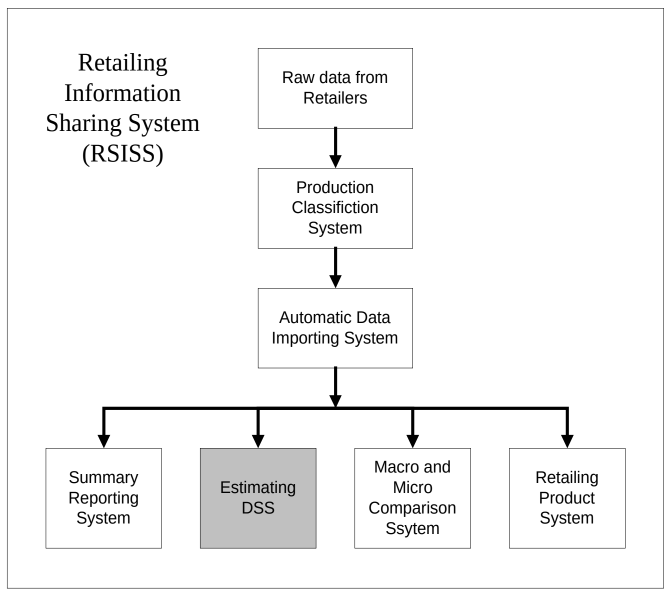

Figure 5: The Position of the Estimating DSS in RSISS

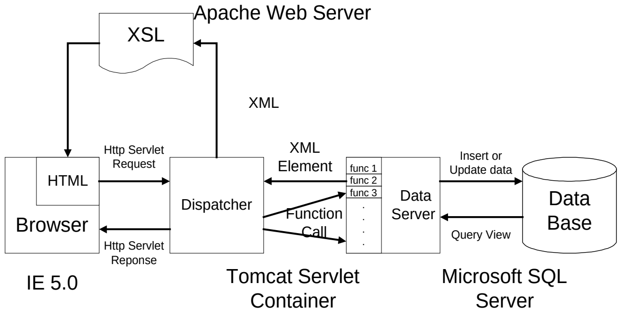

Figure 6: The Architecture of Servers for the Estimating DSS

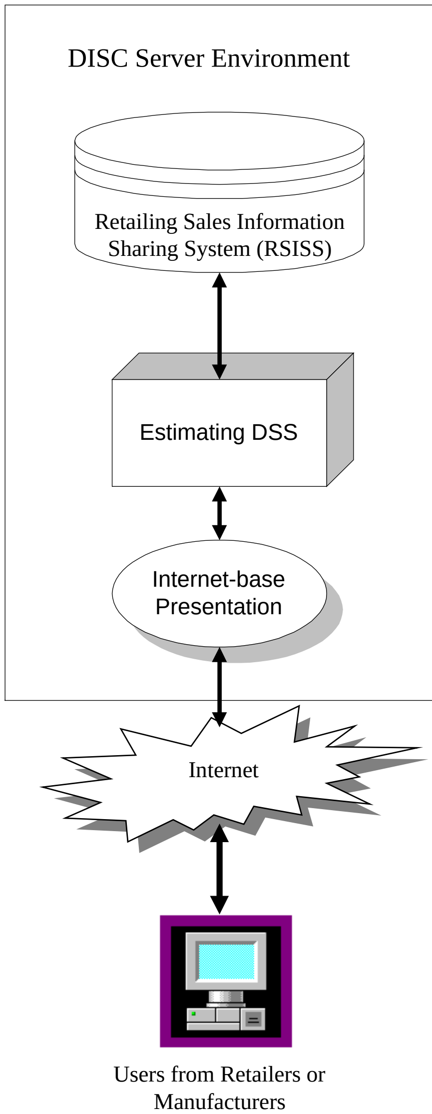

Figure 7: The environment of the Estimating DSS