



PERGAMON

Expert Systems with Applications 25 (2003) 401–411

Expert Systems  
with Applications

[www.elsevier.com/locate/eswa](http://www.elsevier.com/locate/eswa)

# Data mining applied to material acquisition budget allocation for libraries: design and development

C.-H. Wu\*

*National University of Kaohsiung, Information Management, 700, Kaohsiung University Road, Nan-Tzu, Kaohsiung 811, Taiwan, ROC*

## Abstract

Library management frequently faces the need of making the value of the acquired materials significant as far as the most beneficial use of the allocated acquisition budget is concerned. Knowledge in the circulation databases can be explored in-depth to relevantly reflect this need. In this paper, a data mining based model (DMBA) is designed and developed to help allocate the library material acquisition budget by opening up the utilization of library materials that users have made use. The developed model is based on the feature of ID3 algorithm to explore explanatory knowledge via information theory and statistics to derive appropriateness via utilization gain. The main output of the DMBA is the weights as the basis of library material acquisition budget allocation for departments via the combination of explored explanatory knowledge and appropriateness. The developed DMBA was supported by a practical application case.

© 2003 Elsevier Ltd. All rights reserved.

*Keywords:* Acquisition budget allocation; Circulation; Data mining

## 1. Introduction

The objective of the budget allocation is to smoothen the progress of planning and resource allocation in an organization to itemize, slice up and examine all of the products/services that are offered to patrons (Robort, 1998). Seer (2000) has mentioned that in the library budget allocation process, it is important to explain how to arrive at the numbers that are used as a base of budget allocation. To deal with the most advantageous use of materials that are budgeted, it is thought that the performance with respect to the use of materials that an academic department has made should be able to appropriately reflect the budget being allocated. Importantly, since the budget is increasingly limited, it should be an inviolable policy that the more use of materials a department makes, the more budget it is allocated. Consequently, knowledge discovered in the circulation databases is fairly valuable for an academic library to support allocating budget for departments as per appropriateness.

In response to the decision of acquisition budget allocation, Greaves (1974) indicated that the circulation statistics has to be taken into account to reflect the practical demands. The survey provided by Tuten

and Lones (1995) and the study conducted by Budd and Adams (1989) also emphasized that circulation statistics is one of the most extensively referred factors in the decision of desired allocation of budget. Furthermore, in the study of the measurement of the resource utilization efficiency for university libraries, Chen (1997) put the book circulation onto the measurement list as one of the most important dimensions. More recently, Wise and Perushek (2000) have pointed out that circulation is a reliable factor to evaluate the success of material utilization in their research that employed the goal programming as a solution technique to solve the acquisition allocation problem. These research works have indeed shown a substantial aid for the improvement of budget allocation process. However, a drawback that is revealed is the lack of a model that can be employed to help effectively derive the utilization performance with respect to the use of materials in different subjects represented by academic departments, and in consequence cannot pull a proper sign that the operation of budget allocation can rely on. For example, the information that 'the circulation database of this academic year indicated that the department of International Trade made much more use of materials in its subject than others in theirs' is more decision-supportable than that 'the circulation database of this academic year indicated that the department of International Trade made much more use of materials than others' while budgeting.

\* Tel.: +886-75919059; fax: +886-75919328.

E-mail address: [chwu@nuk.edu.tw](mailto:chwu@nuk.edu.tw) (C.-H. Wu).

The techniques that are used to help in determining academic library acquisition budget allocation generally include goal programming and statistics. Goal programming deals with the development of mathematical models that can offer near optimal solutions to the problem of multiple, competitive, and conflicting objectives by giving the rank order of the concerned goals and the constraints of the concerned factors in advance. [Wise and Perushke \(1996, 2000\)](#) have demonstrated the successful use of goal programming for the acquisition allocation problem. The statistics concentrates on the corresponding shared ratio for the concerned factors that are contained in a hierarchical decision tree ([Anderson, Sweeney, & Williams, 1994](#)). Although both techniques have shown a significant contribution to the support of library budget allocation decision for departments, a problem that is seen is the illustration of information discovered in the historical circulation data. In other words, the use of materials that a department has made should appropriately reflect its acquisition budget allocation for the coming academic year. In response to this need, [Kao, Chang, and Lin \(2003\)](#) present a decision support model that can discover information in circulation databases for the library acquisition budget allocation. However, the huge labor consumption for the complex computation may limit its applicability, and thus should be improved from the manual operation toward computerization.

Basically, it is a fairly complex task to deeply explore meaningful information in the historical circulation database that stands for the material utilization for a department. To simply get the percentage of the number of records a department has made use of materials to the total records in the collected circulation database for a period of time is not sufficient. The preprocessing of circulation data, the definition of strength that a material belongs to a department, and the complexity of gain computation are all issues that may cause the task to be fairly complicated. The data collection and preparation needs to gather the circulation data from daily operations and store in a database, clean unnecessary attributes (or fields) and missing data if existing, and reconstruct the created database if necessary. The strength that a material belongs to a department needs to be suitably defined in order for a department to compute the total utilization gain with respect to its performance. For example, the strength that a material is categorized into the subject of financial economics may be defined to be related to the department of Accounting with a degree of ‘absolutely matching’, Information Management with ‘partially matching’, and electrical engineering with ‘not matching’. This implies that when a material classified into financial economics was utilized, the department of Accounting performs better than both Information Management and Mechanical Engineering because it made use of a material in its subject more appropriately. In this case, an assumption that has to be made is that a department uses its own allocated budget to acquire materials only in its subject. In spite of our belief that

some exceptions may occur and are very difficult to eliminate, it is a general policy that the more the suitable use, the more the budget being allocated. For example, although the department of Electrical Engineering may possibly acquire materials in the subject of financial economics, the use made by Electrical Engineering will be regarded as not matching while measuring its performance. In this study, we do not attempt to discourage a department making use of library materials that are not in its subject. However, when the problem of making the value of allocated budget obvious is concerned, we have to consider with greater care whether or not a department makes suitable use of materials.

The Data Mining (DM) with a capability of description and prediction can explore patterns in large historical databases that are meaningful, interpretable, and decision-supportable ([Chen, Han, & Yu, 1996](#); [Fayyad, 1996](#); [Fayyad & Stolorz, 1997](#); [Hirota & Pedrycz, 1999](#)). [Lee and Siau \(2001\)](#) have addressed in a basic detail the use of major DM techniques including artificial intelligence, decision tree paradigm, genetic algorithm, visualization, and statistics. The applications in various domains that are depicted in literature in the past few years have also witnessed the increased use of DM, such as hotel data mart ([Sung and Sang, 1998](#)), personal bankruptcy prediction ([Donato et al., 1999](#)), customer service support ([Hui & Jha, 2000](#)), and the special issue edited by [Kohavi and Provost \(2001\)](#) of a referable journal. [Bigus \(1996\)](#), [Adriaans and Zantinge \(1996\)](#) also presented a fundamental concept for the applicability of DM in business problems covering marketing segmentation, customer ranking, real estate pricing, sales forecasting, customer profiling, and prediction of bid behavior of pilots. In this study, the discovered knowledge with an explanatory form and the measure of appropriateness is used to support the making of the library acquisition budget allocation decision. For example, the explanatory knowledge that ‘the material categories that the department of International Trade made use were more concentrative than those of the other departments’ is helpful for the library budget allocation operation. Moreover, regarding the appropriateness ‘The appropriateness of the use of materials by the department of International Trade was much higher than that of others’ is also fairly decision-supportable.

Usually, the size of a circulation database is very big, and consequently the complexity of gain computation may significantly influence the processing efficiency. This research employs the technique of Structured Query Language (SQL) to overcome this critical problem. The SQL is a well-known database technique that helps in efficiently retrieving the data satisfying a user’s requirements ([Connolly, Beg, & Strachan, 1996](#)). [Han, Fu, Koperski, and Zaiane \(1996\)](#), [Imielinski and Virmani \(1999\)](#), and [Meo, Psaila, and Ceri \(1998\)](#) have successfully made its obvious use to explore knowledge in large databases. Therefore, the SQL is employed in this study to help data preprocessing. The advanced information

متن کامل مقاله

دریافت فوری ←

**ISI**Articles

مرجع مقالات تخصصی ایران

- ✓ امکان دانلود نسخه تمام متن مقالات انگلیسی
- ✓ امکان دانلود نسخه ترجمه شده مقالات
- ✓ پذیرش سفارش ترجمه تخصصی
- ✓ امکان جستجو در آرشیو جامعی از صدها موضوع و هزاران مقاله
- ✓ امکان دانلود رایگان ۲ صفحه اول هر مقاله
- ✓ امکان پرداخت اینترنتی با کلیه کارت های عضو شتاب
- ✓ دانلود فوری مقاله پس از پرداخت آنلاین
- ✓ پشتیبانی کامل خرید با بهره مندی از سیستم هوشمند رهگیری سفارشات