Available online at www.sciencedirect.com

ScienceDirect

Procedia Computer Science 96 (2016) 395 – 402

Procedia Computer Science

20th International Conference on Knowledge Based and Intelligent Information and Engineering Systems, KES2016, 5-7 September 2016, York, United Kingdom

Predicting Book Use in University Libraries by Synchronous Obsolescence

Kensuke Baba[a,*], Toshiro Minami[b], Tetsuya Nakatoh[c]

[a] *Fujitsu Laboratories, Kawasaki, 211-8588, Japan*
[b] *Kyushu Institute of Information Sciences, Dazaifu, Fukuoka, 818-0117, Japan*
[c] *Kyushu University, Fukuoka, 812-8581, Japan*

Abstract

Collection management in libraries should be conducted on the basis of analyses of circulation data. In addition to the number of times a book is loaned out, the change in the number of loans over time can be a useful measure for managing books. In this study, we investigate the relationship between diachronous obsolescence (that is, the change in use over time) and synchronous obsolescence (that is, the distribution of use over the elapsed time from accession, for a relatively short period). We evaluate three algorithms based on the two types of obsolescence, in order to predict future book use from past use information. We apply these algorithms to actual circulation data from a university library and investigate prediction accuracy. The results show that the synchronous obsolescence-based algorithm can predict future book use more accurately than the diachronous obsolescence-based algorithms, especially for information recorded over a relatively short period.

Peer-review under responsibility of KES International

Keywords: Data mining; library; circulation data; obsolescence model; machine learning.

1. Introduction

Collection management is an important operation for libraries. In typical university libraries in Japan, collection management processes such as selection, arrangement, and disposal of books are being conducted according to static rules or the intuitions of the library staff. To meet the constantly changing demands of library users, these processes should be improved on the basis of the actual use of books in libraries.

Some methods to improve collection management processes can be reduced to the problem of predicting future book use from past use information. For a given set of books and past circulation data, prediction problem aims to predict the number of loans for the books in the set by analyzing the circulation data. For example, Silverstein and Shieber[1] solved the problem for selecting books to be stored in an off-site location. This work was applied to the arrangement of books, and the idea can be extended to the disposal of books. As for the selection of books, the idea

* Corresponding author. Tel.: +81-044-754-2328 ; fax: +81-044-754-2771.
E-mail address: baba.kensuke@jp.fujitsu.com

Peer-review under responsibility of KES International
doi:10.1016/j.procs.2016.08.082

of a "good" selection can be defined using the frequency of use for the selected books. Evans[2] investigated the use of books classified by the person who selected the books in several academic libraries. Pritchard[3] investigated the use of books classified by the manner of selection. Koizumi[4] also investigated the relationship between book use and the manner of book selection at the Keio University library in Japan. These investigations of the number of loans in various cases of selection are applicable to the problem, even if the future number was not predicated explicitly.

An effective approach to the prediction problem is using the change in the number of loans over time, while the work described in the previous paragraph is based on classification of loan records according to various attributes. In the work by Silverstein and Shieber, various methods that analyze circulation data for the problem were combined on the basis of a case analysis using decision trees, which were constructed using bibliographic information such as the date of publication and the language. It is experientially known that the number of loans for a book (or a set of books) decreases over time. This decrease is called "(diachronous) obsolescence[5]." Matsui and Isono[6] confirmed a diachronous obsolescence in book use in the Nara University library in Japan. Diachronous obsolescence can be approximated by a "synchronous[5]" obsolescence, that is, the distribution of use with the "oldness" of the books over a short period. The oldness of a book is usually defined as the elapsed time since the publication or the accession of the book to the library. The relationship between diachronous and synchronous obsolescence has been mentioned in citations of scholarly papers in various work[5,7]; however, for books in libraries, this relationship has not been clarified through an analysis of actual circulation data.

In this study, we evaluated three prediction problem algorithms, in order to clarify the difference between diachronous and synchronous obsolescence. Two of the algorithms are based on diachronous obsolescence, and the other is based on synchronous obsolescence. The three algorithms use linear regression to estimate the change in the number of book loans over time. The diachronous obsolescence-based algorithms use a function to estimate the change in the number of loans for the target book set (based on past data), and then predict the number of future loans using the obtained function. To perform the estimations, one of the algorithms uses a linear function, while the other uses an exponential function. Clearly, these algorithms require past data from a sufficiently long period to predict future use accurately. The algorithm based on synchronous obsolescence estimates the distribution of the number of loans over the elapsed time, and then uses the obtained parameter to create a model for the estimation. This algorithm can use past information for other books of the target book set, and hence is expected to be suitable for prediction in case with insufficient past information.

We applied the three algorithms to actual circulation data and evaluated their suitability in terms of prediction accuracy. We used loan records obtained from the Kyushu University library in Japan, for the period from December 2009 to November 2014. First, we predicted the number of loans for books classified by the year they were accessioned. The evaluation was conducted with various subsets of circulation data, to clarify the effect of the period of data available for prediction. Next, to evaluate the suitability of the algorithms for restricted sets of books, we predicted the number of loans for books classified by subject. The effects that the characteristics of subjects had on obsolescence have been investigated for journal citations[8] and for library books in our previous work[9].

Our study will suggest a viewpoint for improving the collection management in university libraries. According to the results of the evaluation, we can estimate the demands for books more accurately; this should lead to better decisions in the selection, arrangement, and disposal of books in libraries.

The remainder of this paper is organized as follows. Section 2 introduces three algorithms designed to predict the number of book loans on the basis of circulation data; it then describes the collected circulation data and the method to evaluate the algorithms with the data. Section 3 reports the results of the evaluation. Section 4 provides an analysis of the results, and discusses future directions of our study.

2. Methods

The purpose of our study is to evaluate the possibility of substituting synchronous obsolescence for diachronous obsolescence to predict the use of library books. We defined three algorithms that use past circulation data to predict the number of times a book will be loaned out in the future. Two are standard algorithms that analyze the change in the number of loans over time. The other is based on the distribution of the number of loans over the elapsed time from accession. We apply the algorithms to actual circulation data from a university library to evaluate the algorithms in terms of prediction accuracy.