AJIT-e: Online Academic Journal of Information Technology
2018 Yaz/Summer – Cilt/Vol: 9 - Sayı/Num: 33
DOI: 10.5824/1309-1581.2018.3.005.x
http://www.ajit-e.org/?menu=pages&p=details_of_article&id=358

Received: 06.07.2018 **Editorial Process Begin: 30.07.2018** **Published: 05.10.2018**

Improving Service Processes and The Library Quality Service with Data Mining Methods

Yüksel YURTAY, *Sakarya University, Computer Engineering Department, Sakarya, Turkey, yyurtay@sakarya.edu.tr, ORCID: 0000-0003-1814-3432*
Özgür ÇİFTÇİ, *Sakarya University, Computer Engineering Department, Sakarya, Turkey, ociftci@sakarya.edu.tr, ORCID: 0000-0002-1308-7585*
Eyüp AKÇETİN, *Muğla Sıtkı Koçman University, Muğla, Turkey, eyup.akcetin@mu.edu.tr, ORCID: 0000-0001-7232-2154*

Özet İçinde bulunduğumuz dijital çağda tüketicileri bilgilerini toplayan İşletmeler; müşterilerine daha iyi hizmet vermek ya da yeni pazar alanları açabilmek adına tüketici bilgilerinden çıkarımlar yapmaktadır. Bu işlemlerden sonuçlar elde edebilmek için de yapay zeka teknikleri ile birlikte veri madenciliği yöntem ve algoritmaları kullanmaktadırlar. Yoğun rekabet ortamında veri madenciliği yöntemleri hizmet üreten süreçlerde de etkin bir biçimde kullanılmaktadır. Müşteri ilişkileri bağlamında bu teknikler kütüphane süreçlerinin iyileştirilmesinde ve hizmet kalitesinin artırılmasında veri madenciliği yöntemlerinden yararlanılmaktadır. Bu çalışma kapsamında İlk olarak kullanıcı profiline ve arama geçmişine dayalı kümelemeler gerçekleştirilmiştir. İkinci aşamasında kümeler içerisindeki okuyucu taleplerindeki frekanslar belirlenerek kütüphane hizmetlerinin yapılanmasına dönük değerlendirmeler yapılmıştır. Bu çalışmanın veri setini belirli bir zaman aralığında bir kamu üniversitesinin veri tabanından alınan veriler oluşturmaktadır. Çalışmada yöntem olarak söz konusu kütüphanenin verilerine veri madenciliği yöntemlerinden kümeleme ve apriori algoritması kullanılmıştır. Sonuç olarak kütüphane verileri üzerinde uygulanan veri madenciliği algoritmaları ile kütüphanede verilen hizmet süreçlerinin ve bu hizmetlerin kalitesinin iyileştirilmesi için okuyucu profillerinin tanımlanması, dönemsel değişen okuyucu taleplerinin görüntülenmesi ve okuyucuların kütüphane etkileşimleri belirlenmiştir. Sonraki çalışmalarda farklı üniversitelerden kütüphane verileri alınarak daha farklı araştırmalar, kıyaslamalar yapılabilir ve farklı sonuçlar elde edilebilir. Aynı zamanda farklı kütüphanelerin okuyucuları için çalışmalar tekrarlanarak sonuçlar değerlendirilebilir. Benzer çalışmalarda, veri miktarı ve zaman aralığının, sonuçlar üzerinde etkili olabilir.
Anahtar Kelimeler: Veri Madenciliği, Veri işleme, Süreç iyileştirme, Kütüphane

Abstract In this digital age, businesses that collect consumer information; provide better service to its customers or new market areas in the name of consumer information from the open mining, artificial intelligence techniques, methods and algorithms in conjunction with data mining are used in order to be able to get the results of these operations.Data mining methods in an intensive competition environment in the service-producing process, are being used effectively. For customer relations in the context of these techniques in improving library processes and in improving the quality of service data mining method is utilized.The data aim-based of this study was to set a specific time range of the data that is received from a public University. Study

DOI: 10.5824/1309-1581.2018.3.005.x

of the library data in question required data mining methods of clustering and the apriori algorithm.As a result, the library data are implemented on data mining algorithms in the library with the given service processes and improvement of the quality of these services being the defining periodic changing of profile reader to reader demands and determined display and readers library interactions. Subsequent studies based on data from different universities library may lead too to different research, comparisons, and therefore different results. At the same time, the results of repeated studies for the readers of different library can be evaluated. İn similar studies, the amount of data and the range of time can be effective on the results too.

Keywords: Data mining, data processing, process improvement, Library

1. Introduction

For the solutions in our everyday life, the need to utilize the digital data which spur and continue to grow and grow has been effectively compulsory. This necessity has driven us to produce solutions in all areas of our lives. Data mining and statistical analysis methods are also called "information discovery in databases". In a nutshell, data mining is the task of gaining valuable knowledge from large-scale data. In this regard, it is possible to reveal the relationships between the data and to make forward-looking predictions when necessary (Özkan, 2008). "Data Mining" includes this phase of model building and evaluation.

Data Mining is a search for relations and rules using computer programs that will allow us to generate predictions about the future from a large amount of data (Erpolat, 2012). In times of increasing competition, we collect information from our customers; They use data mining methods and artificial intelligence techniques to see changing needs, and differentiated markets. The methods adopted are also used effectively in service producing sectors. In the context of customer relationships, it is benefiting from improvements in processes that have a similar operating structure from markets and markets. Libraries are one of the institutions that have such structure. Libraries, today, are continuing to see intense demand. As a result of our interviews with our observations and library staff, it has shown that the density is about access to resources and that improvements can be made. Reducing the intensity, facilitating access to resources will foster an increase in the quality of libraries service. In order to improve the library services, various studies have been carried out but the results obtained have not been evaluated in the application stage. In our study, the most preferred resources will be determined in the library environment by analyzing the association (fig.1). Then, readers' requests in terms of access to resources, will be evaluated by library operation and suggestions for the locations of the resources will be created. As a result, we will improve the quality of service and re-evaluate the library's resource allocation plan by making it easier for readers to change their access to resources. In other parts of our work, we have researched these problems and shared the data mining algorithm we use.

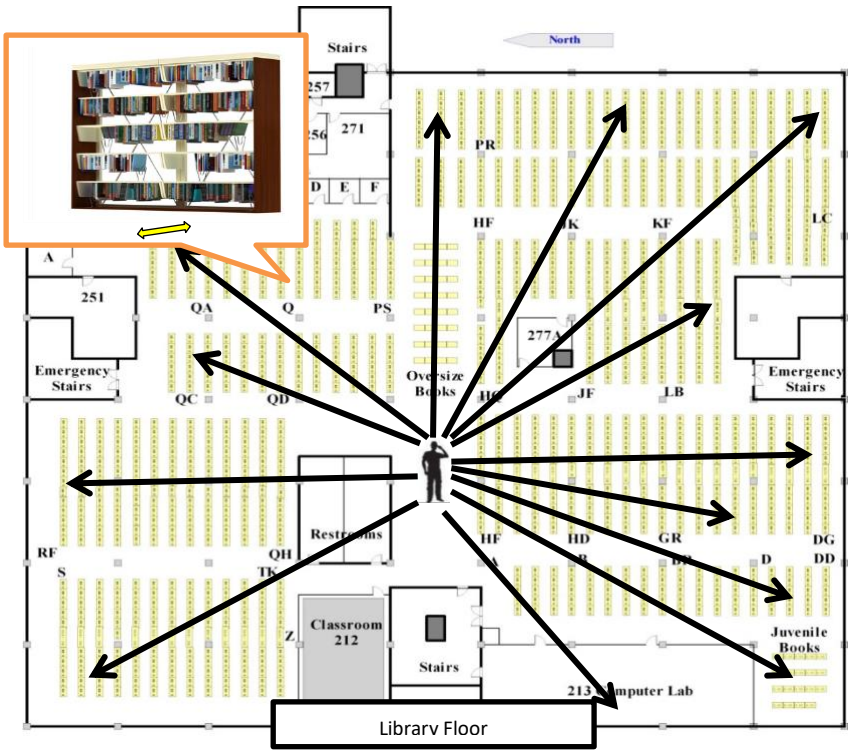

Figure 1. Current library resource access plan.

1.1. Previous Studies

In our study, data mining techniques based on the cooperativeness analysis were used and the rules of association with the a priori algorithm were consequently obtained. There is not much research in the literature about a priori and cohesion analysis. By using the a priori algorithm and the FP-Growth Algorithm, Wei Z., Hongzhi L. and Na Z. (2008) solution have obtained two known problems. For mobile service users Surendiran R., Rajan.K.P. ve Sathish K.M. (2010), came up with the rules of FP-Growth Algorithm. Moving from the data in the market database Vishal S., Nikita J., Shahid V. (2010), have drawn up the rules of association for products sold using the a priori algorithm. Erpolat Sema (2012) compares the A priori and FP-Growth algorithms in determining the co-existence rules for both the algorithm and the application instance in automobile authorized services.

Conclusions have reached the observations that the FP-Growth Algorithm is consistent and quicker than the A priori Algorithm. Translation errorIn this study, the A priori Algorithm, reader data from a university library were used to improve service processes and quality of service, identification of reader profiles, periodic changes in reader demands, and reader interactions with libraries.

1.2. Data mining

Data mining is the process by which a dynamic process is obtained from data chunks of valid, practicable information, which is simply, unclear and unpredictable from the data at hand. In other words, it allows large quantities of data to be analyzed to discover meaningful patterns and rules (Berry and Linoff, 2004, p.12). As digital data begin to accumulate in the electronic environment, data mining has widely become used-tool in every field. It is a technique which has especially gained importance in recent years both in the world and in Turkey. It is mainly studied in two main sections. The first is a descriptor that is used to

DOI: 10.5824/1309-1581.2018.3.005.x

estimate unknown data, and the other is a descriptor that identifies the data at hand (Akpınar, 1998, p.5).

The most frequently used models are the classification (estimator), clustering analysis (descriptive), association rules and regression analysis.

1.3. Rules of Association

It is among the descriptive models of data mining. The rules of association are intended to divulge the interrelated data and to determine the magnitude of the link between them. Association analysis is defined as the discovery of associative rules of property values that occur at high frequency in a particular data set (Argüden and Erşahin, 2008). For association analysis, many algorithms that provide information from large data sets can be categorized sequentially or in parallel. Sequential algorithms include logical expressions in which product clusters are constructed and counted. In the study, a priori algorithm is investigated from successive algorithms.

1.4. Apriori Algorithm

The a priori is derived from the consecutive algorithms in the rule-ordering of association, derived from the word "prior". It has a recursive nature and is used in the discovery of data sets that pass frequently in the data stacks.

The algorithm consists of the following steps;

a) Threshold values are first determined to compare the support and trust criteria in order to be able to perform the association analysis. Translation errorIt is expected that the results obtained from the application will be equal to or greater than these threshold values.
b) For each product to be included in the analysis by scanning the database, the recounts, i.e. support numbers, are calculated. This number of supports is compared to the number of threshold supports. Lines with small numbers of threshold support are removed from the parser and the appropriate records are taken into account.
c) The products selected in the above step are grouped in pairs to obtain the number of repetitions of these groups. These numbers are compared to the threshold support numbers. Lines with lower values than the threshold value are removed from the parsing.
d) This time, threes, quarters, etc. grouping is performed to obtain the support numbers of these groups and compared with the threshold values, the processes are continued as long as the threshold values are appropriate.
e) After the product group is identified, the rules of cohesion are derived from the rule support scale and confidence measures are calculated for each of these rules. (Ozkan, 2008).

To establish relationships in the association analysis, two criteria “support” and “trust” are used. While the "support criterion" specifies the order in which a relationship is replicated across all exchanges, the "trust criterion" specifies the probability that the selection of an X group of objects will make the choice of Y groups of objects.

The combination of choosing the object X and the object Y is expressed in the form of rule X → Y, and the rule is expressed in the form of support criterion.

$$\textbf{Support } (X \rightarrow Y) = \textbf{number } (X, Y) / N$$

Here, the number (X, Y) represents the number of supports containing the group of objects X and Y together.
N is the total number of choices. The trust criterion that expresses the probability that the X and Y object groups are selected together.

Trust (X → Y) = number (X, Y) / number (X).

The relationship between objects is calculated by means of support and trust. Translation errorThe high level of support and confidence of the two objects is an indication that the association is also important.

2. Methodogy

This work was based on the CRISP-DM Methodology developed by Daimler Chrysler, SPSS, NCR and OHRA and accepted in the data mining literature. Translation errorAccording to the CRISP-DM methodology, the data mining process consists of six stages:

1) Job understanding,
2) Data understanding,
3) Data preparation,
4) Modelling,
5) Evaluation,
6) Application (Chapman, et al., 2000).

2.1. Scope

The aim of the study is to provide a measurable quality service increase and reorganization of the library system by improving access to resources in library reader services, data mining techniques.
We have witnessed the changing services being questioned and reorganized in our surrounding. Libraries are the most important institutions where information is available, and in this sense, they are the most important institutions that should provide service-access to all researchers and readers. Thus, the library service with changing lifestyles and needs, questioning their services, it has become inevitable. Issues such as accessibility in the library, positioning of the shelf, reader classification and resource management have emerged as processes that need to be restructured. According to the order of the most demanded books and subjects, the determination of library shelf order has become important in terms of accessibility. The extraction of various statistics according to the reader profiles and the request, the evaluation and diversification of the service provided has become very important. Improved access to library resources through technological communication tools and data processing techniques in this context will also be a separate achievement. Data mining techniques applied to organizations with similar layouts are effectively being implemented. It will provide effective use of institutional resources through business intelligence approaches and data mining techniques on data accumulated within the organization. In this frame, the data mining algorithm like a priori, is an important tool in performing association analysis. Today, product placement patterns of institutional shopping centers are evaluated according to the results of the a priori algorithm. Thus, the accessibility and demand quantities of the products are increased and the costs of possession, storage and transportation are minimized. Similarly, library functioning can be evaluated in this framework. When the study is modeled and completed, the changes in the reader profile will quickly identify the changes to be made in the book / publication layout and steer suggestions for rack configuration. Such proposals will not only provide the fast and quality

DOI: 10.5824/1309-1581.2018.3.005.x

service that the readers expect from the library, but will also facilitate the process of the library staff as well as saving work force and space.

2.2. Data Preparation and Analysis

In the study, a university library was solved using library reader request information and library settlement information. The existing reader data is converted into processable data through the following

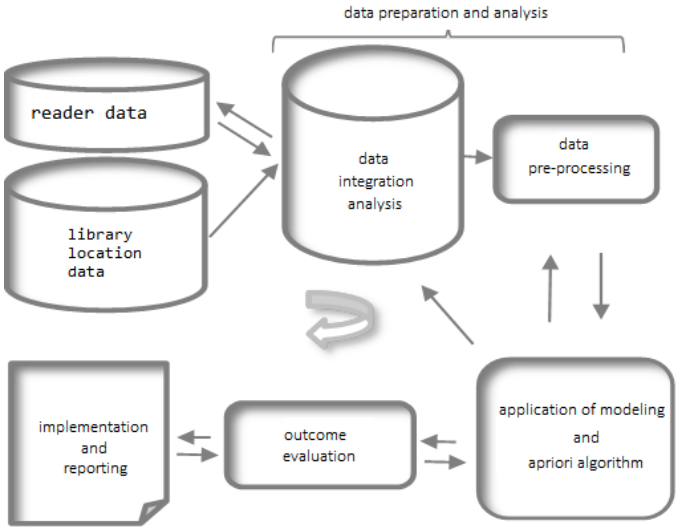

Figure 2. Scope of Work

processes. After the data preprocessing phase, the model is installed and the application of the algorithm is completed. The results are evaluated by how they can be applied to the library operation processes. In the realization phase, the evaluation and the results are reflected in the changes in the library book access processes (Figure 2).

2.3. Modeling and Application

Approximately 122,000 data sets have been implemented so that the model can be pre-processed. After testing with sample data sets, the entire data set can be assembled by modeling the frequently requested books together.

The model was applied on RapidMiner (Figure 2).

Row No.	Yazar	Baslik	Konu
7	Ülken, Hilmi Ziya.	Türkiye'de çağdas düsünce tarihi / Hilmi Ziya Ülken.	Islam and state Turkey.
8	Ülken, Hilmi Ziya.	Türkiye'de çağdas düsünce tarihi / Hilmi Ziya Ülken.	Islam and state Turkey.
9	Ülken, Hilmi Ziya.	Türkiye'de çağdas düsünce tarihi / Hilmi Ziya Ülken.	Islam and state Turkey.
10	Ülken, Hilmi Ziya.	Türkiye'de çağdas düsünce tarihi / Hilmi Ziya Ülken.	Islam and state Turkey.
11	Ülken, Hilmi Ziya.	Türkiye'de çağdas düsünce tarihi / Hilmi Ziya Ülken.	Islam and state Turkey.
12	Ülken, Hilmi Ziya.	Türkiye'de çağdas düsünce tarihi / Hilmi Ziya Ülken.	Islam and state Turkey.
13	Ülken, Hilmi Ziya.	Türkiye'de çağdas düsünce tarihi / Hilmi Ziya Ülken.	Islam and state Turkey.
14	Ülken, Hilmi Ziya.	Türkiye'de çağdas düsünce tarihi / Hilmi Ziya Ülken.	Islam and state Turkey.
15	Ülken, Hilmi Ziya.	Türkiye'de çağdas düsünce tarihi / Hilmi Ziya Ülken.	Islam and state Turkey.
16	Ülken, Hilmi Ziya.	Türkiye'de çağdas düsünce tarihi / Hilmi Ziya Ülken.	Islam and state Turkey.
17	Ülken, Hilmi Ziya.	Türkiye'de çağdas düsünce tarihi / Hilmi Ziya Ülken.	Islam and state Turkey.
18	Ülken, Hilmi Ziya.	Türkiye'de çağdas düsünce tarihi / Hilmi Ziya Ülken.	Islam and state Turkey.
19	Ziya Gökalp, 1876-1924	Türkçülüğün esasları / Ziya Gökalp ; hazırlayan: Kemal Bek.	Milliyetçilik Türkiye.
20	Ziya Gökalp, 1876-1924	Türkçülüğün esasları / Ziya Gökalp ; hazırlayan: Kemal Bek.	Milliyetçilik Türkiye.

Figure 3. Example of processed data.

Stage 1: Assignment and identification given.
Stage 2: Transformations between data.
Step 3: Converting to available rules, the process is running.

Figure 4 shows the Rapidminer model.

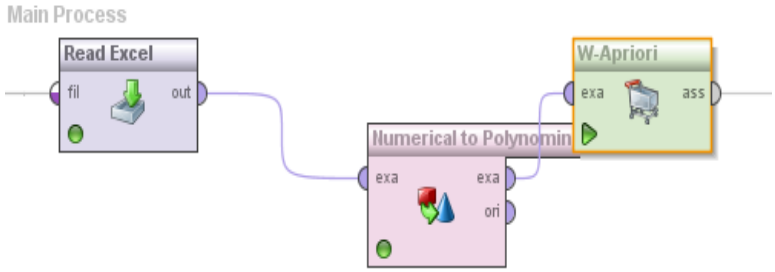

Figure 4. Rapidminer Model

Obtained associations and conclusions (Figure 5) shows a sample demand figure for book names and authors.

DOI: 10.5824/1309-1581.2018.3.005.x

Figure 5. Association Rules

When the sample report in Figure 5 is examined, the most preferred author is "Ziya Gökalp"; the most read book is "Ziya Gökalp's principles of Turkism", and the most repeated book topic is "Nationalism Turkey".

In order to make it workable data, definitions were made so that the Apriori algorithm could be applied on the RapidMiner Application after cleaning, correcting and normalizing the incorrect entries of the employees and the result was reported.

3. Conclusion

In this study, the RapidMiner data mining program was used and association rules were derived through the a priori algorithm. Using data from a university library, readers' associations were found in book requests. The current settlement order of the library and the rules of association obtained are evaluated and evaluated, and recommendations on library functioning books and shelves are put forward. The emerging order has facilitated and accelerated access to library resources. The new access scheme is shared in Figure 6. Created references are shared with library administrators. The process applied in our work can be repeated according to changing reader requirements. If the algorithm of the study is coded

by any software language, it is evaluated that the settlement order can be queried in certain periods and settled to a dynamic process.

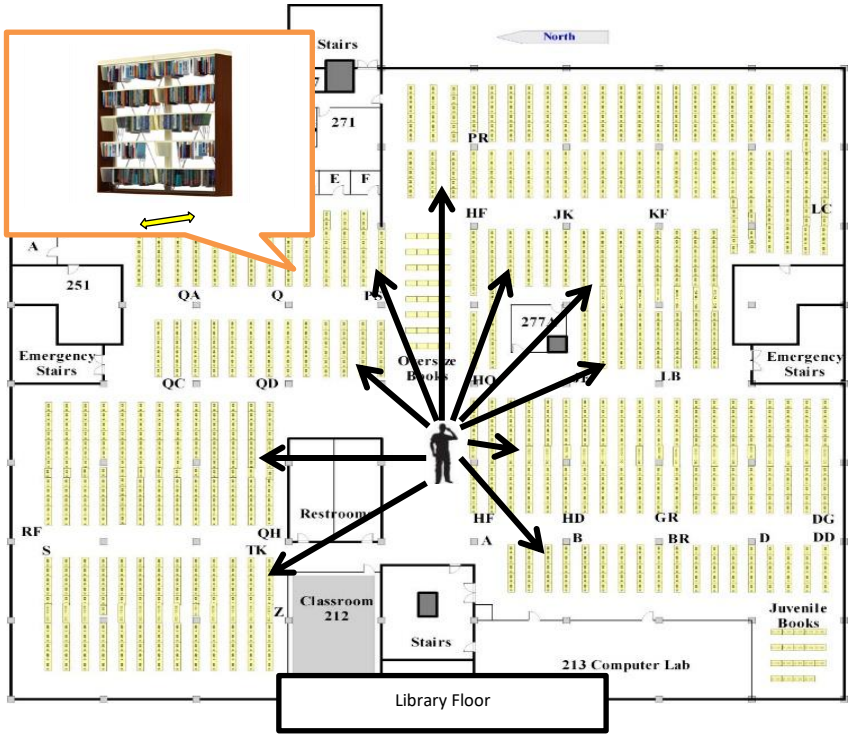

Figure 6. Example of recommended library access scheme.

4. References

Akpınar, H. Veri Tabanlarında Bilgi Keşfi ve Veri Madenciliği, İ.Ü. İşletme Fakültesi Dergisi. C:29, 1-22 ss; 2000.

Argüden, Y., Erşahin, B. Veri Madenciliği: Veriden Bilgiye, Masraftan Değere,1.Basım, ARGE Danışmanlık Yayınları, İstanbul; 2008.

Berry, M.J.A ve Linoff, G.S, Data Mining Techniques: For Marketing, Sales and Customer Relationship Management, New York: John Wiley & Sons Inc; 2004.

Chapman, P. CRISP-DM 1.0 Step-by-step data mining guide, SPSS Inc; 2000.

Erpolat Semra, Otomobil Yetkili Servislerinde Birliktelik Kurallarının Belirlenmesinde Apriori ve FP-Growth Algoritmalarının Karşılaştırılması, Sosyal Bilimler Dergisi(C12S2), Anadolu Üniv; 2012.

Özkan, Y. Veri Madenciliği Yöntemleri, 1. Basım, Papatya Yayınları; 2008.

Surendiran R., Rajan.K.P. ve Sathish K.M., Study on the Customer Targeting Using Association Rule Mining. Surendiran et. al., (IJCSE) International Journal on Computer Science and Engineering, Madurai, India, Vol. 02, No. 07, 2010, 2483-2484, ISSN : 0975-3397; 2010.

Vishal S., Nikita J. ve Sharad V., (). Efficient Use of Apriori Algorithm for Accesing Supermarket Database. IT and Business Intelligence. Proceedings of 2nd international Conference on IT & Business Intelligence (ITBI-10), India, Technically Sponsored by IEEE CIS ISBN No: 978-81-7446-900-7; 2010.

Wei Z., Hongzhi L. ve Na Z., Research on the FP Growth Algorithm about Association Rule Mining, Business and Information Management, ISBIM '08. International Seminar on, pp. 315-318, ISBN: 978-0-7695-3560-9; 2008.

Zhang W., Liao H., Zhao N., Research on the FP Growth Algorithm about Association Rule Mining. 2008 International Seminar on Business and Information Management, Hong Kong, pp. 315-318, ISBN: 978-0-7695-3560-9; 2008.