Association Analysis of Course Data In Online Education Systems

Denizay Biberoğlu, Yusuf Yemliha Çelik

Bilgisayar Mühendisliği Bölümü

Yıldız Teknik Üniversitesi, 34220 Istanbul, Türkiye

{denizay.biberoglu, yemliha.celik}@std.yildiz.edu.tr

Özetçe —**Birliktelik analizi, verilerin birbirleriyle olan ilişkilerinin analiz edilmesi ve birliktelik kuralları çıkarılmasına denir. Birliktelik analizi gerçek hayatta günümüzde birçok şirket tarafından kullanılmaktadır. Müzik dinlediğimiz uygulama, alışveriş yaptığımız e-ticaret siteleri gibi birçok şirket bize yeni ürün önermek için ellerinde bulundurdukları veri üzerinde birliktelik analizi uygulayarak hangi ürünlerin birlikte tüketildiği bilgisini çıkarmaktadır.**

Proje kapsamında UZEM verilerinden öğrencilerin seçmiş oldukları dersler üzerinde birliktelik analizi yapılmış ve kayıtlar üzerinden belirli istatistiksel veriler elde edilmiştir. Proje süresince, başlangıçta veri incelenmiş, gereksiz veriler temizlenmiş ve gerekli olduğu durumlarda yeni veriler oluşturulmuştur. Daha sonra istenilen senaryolar için veri düzenlemiş ve belirli yazılım kütüphaneleri aracılığıyla birliktelik kuralları ve istatistiksel veriler oluşturulmuştur.

Anahtar Kelimeler—*birliktelik analizi, apriori, veri madenciliği.*

Abstract—**Association analysis is the analysis of the relationships between data and the derivation of association rules. Association analysis is used by many companies in real life today. Many companies such as the application we listen to music, e-commerce sites where we shop, etc. apply association analysis on the data they have in order to recommend new products to us and extract information about which products are consumed together.**

Within the scope of the project, association analysis was performed on the courses selected by students from UZEM data and certain statistical data were obtained from the records. During the project, the data was initially analyzed, unnecessary data was cleaned and new data was created when necessary. Then, the data was organized for the desired scenarios and association rules and statistical data were created through specific software libraries.

Keywords—association analysis, apriori, data mining.

I. Introduction

Data mining is the process of analyzing data from different angles and summarizing it into useful information. It allows users to analyze and categorize data and summarize relationships between data. Technically, data mining is the process of finding correlations or patterns in large relational databases. It includes some common tasks such as anomaly detection, clustering, association analysis, regression, summarization, classification, etc [1]. Within the scope of the project, association analysis, one of the data mining methods, will be used.

Association analysis aims to extract interesting correlations, frequent patterns, relationships or coincidental structures between sets of items in databases. Association rules are widely used in various fields such as telecommunication networks, market and risk management, inventory control, etc.

Association analysis is to find association rules that provide a predefined minimum of support and confidence from a given database. The problem is usually decomposed into two sub-problems. The first is to find item sets in the database that are above a predefined threshold; these item sets are called frequent or large item sets. The second problem is to generate association rules with minimum confidence constraints from these large item sets [2].

II. Related Work

Although association analysis with the Apriori algorithm has been applied in many areas before, it has almost never been used in the field of course recordings, which is within the scope of our project.

III. Overview

Apriori algorithm first creates frequent item sets for given data set. When the support value of an item set exceeds a certain threshold, it is considered a frequent item set. This threshold often called Minimum Support. Frequent item sets created with given steps:

- Create a list of all the elements that appear in every transaction and create a frequency table.

- Set the minimum level of support and eliminate items with lower level of support than minimum support.

- Create all potential pairings of important elements. Bear in mind AB and BA are same.

- Count the number of times each pair appears in a transaction.

- Eliminate pairs with support value below minimum support.

- To create three-item set, pairings with the same initial letter can be joined together. In example, consider you have pairings OP, OB, PB and PM. You can have OPB and PMB item sets from those pairings.

- Repeat until you end can not create any more frequent item sets [3].

After frequent item sets are created, association rules can be created as well. Association rules in Apriori algorithm created in the following steps:

- Set the minimum support and confidence values.

- Take all the transaction's supports that are greater than the chosen minimum support value.

- Create the rules with the frequent item sets we have.

IV. METHODS

A. Data Preprocessing

Before creating frequent item sets and association rules, we need to preprocess and transform our data.

We removed unnecessary rows and extracted needed data. We needed vocational elective course and social elective course data for our analysis. So we extracted those rows to use in frequent item set and association rule generation.

B. Frequent Item Set and Association Rule Generation

To create association rules, first we needed to create frequent item sets. We set the minimum support value for frequent item sets to %2. Then we fed the frequent item sets to Apriori algorithm to create association rules for each department.

After creating rules for each department, we noticed there are lack of rules for social selective courses in some departments. We determined that is cause of this is we did not have sufficient data.

Antecedents	Consequents	Antec. Sup.	Conseq. Sup.	Support	Confidence	Count
Uzman Sistemlere Giriş	Veri Madenciliğine Giriş	0.15	0.45	0.08	0.54	28
Uzman Sistemlere Giriş	Yapay Sinir Ağlarına Giriş	0.15	0.22	0.05	0.31	16
Uzman Sistemlere Giriş	Yapay Zeka	0.15	0.22	0.04	0.27	14

Figure 1 Top three rules for class "Introduction to Expert Systems". It can be seen that class' top 3 rules are related to the field artificial intelligence, which is class itself is related to as well.

C. Statistical Analysis

We decided to create 2 statistical data from our given data. First, we extracted which courses were taken most in each department. Second, we extracted which other department students take classes from other than their primary department.

This two statistical data and association rules are both can be viewed from GUI application.
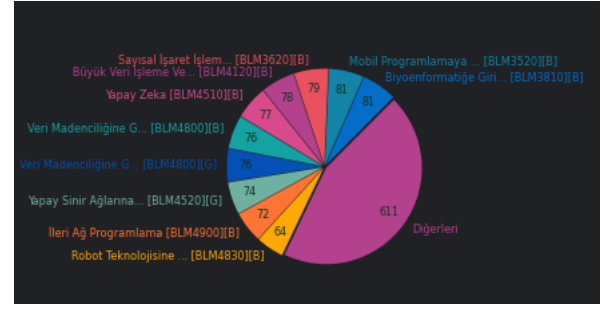
Figure 2 Most selected classes in department "Computer Engineering" that can be viewed from GUI.
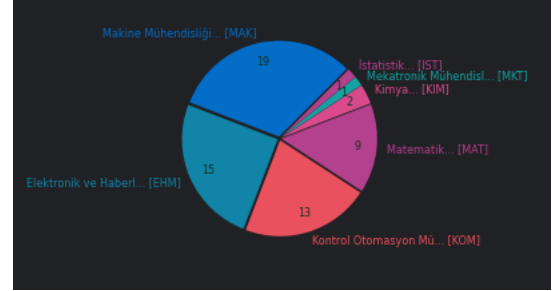
Figure 3 Pie chart that shows which departments that students from "Computer Engineering" take classes from.

V. CONCLUSION

Association analysis can be used in many fields and it creates rules by establishing relationships between seemingly independent events.

In our project, we decided to use the Apriori algorithm to perform association analysis. At the end of our work, we ended up with an application with GUI that user can view some statistical data and association rules for desired departments or classes. It is sufficient to give advice to students in the field of education.

The Apriori algorithm was not able to generate enough rules for each department or each course because the data we had was not sufficient. To overcome this problem, we need to collect data for a longer period of time and re-analyze and data mine this data.

REFERENCES

[1] T. A. Kumbhare and S. V. Chobe, "An overview of association rule mining algorithms," *International Journal of Computer Science and Information Technologies*, vol. 5, no. 1, pp. 927–930, 2014.

[2] S. Kotsiantis and D. Kanellopoulos, "Association rules mining: A recent overview," *GESTS International Transactions on Computer Science and Engineering*, vol. 32, no. 1, pp. 71–82, 2006.

[3] M. Banoula. (2023) What is apriori algorithm in data mining: Everything you need to know. [Online]. Available: https://www.simplilearn.com/tutorials/machine-learning-tutorial/apriori-algorithm-in-data-mining