ФЕДЕРАЛЬНОЕ ГОСУДАРСТВЕННОЕ БЮДЖЕТНОЕ ОБРАЗОВАТЕЛЬНОЕ УЧРЕЖДЕНИЕ ВЫСШЕГО ОБРАЗОВАНИЯ

«НОВОСИБИРСКИЙ ГОСУДАРСТВЕННЫЙ ТЕХНИЧЕСКИЙ УНИВЕРСИТЕТ»

КАФЕДРА ВТ

ОТЧЁТ ПО ЛАБОРАТОРНОЙ РАБОТЕ №3

«Ассоциативный анализ» по дисциплине «Компьютерные технологии анализа и обработки данных»

Выполнили: студенты

гр. АММ2-24

Атласюк Игорь Романович

Ириков Евгений Алексеевич

Проверил: к.т.н., доцент Кафедры ВТ Альсова Ольга Константиновна

Новосибирск 2024

## Содержание

[Постановка задачи 3](#_bookmark0)

[Ход работы 4](#_bookmark1)

[Заключение 23](#_bookmark2)

[Приложение 24](#_bookmark3)

# **Постановка задачи**

Изучить возможности модуля «Ассоциативный анализ» среды R на

примере решения прикладной задачи.

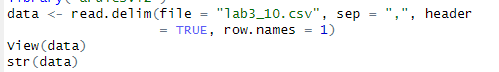
# **Ход работы**

Для выполнения ассоциативного анализа в среду необходимо подгрузить два пакета:

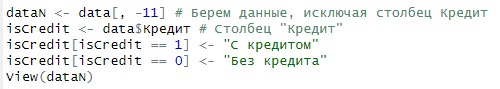
* arules – пакет для проведения ассоциативного анализа;
* arulesViz – пакет для визуализации результатов ассоциативного анализа.



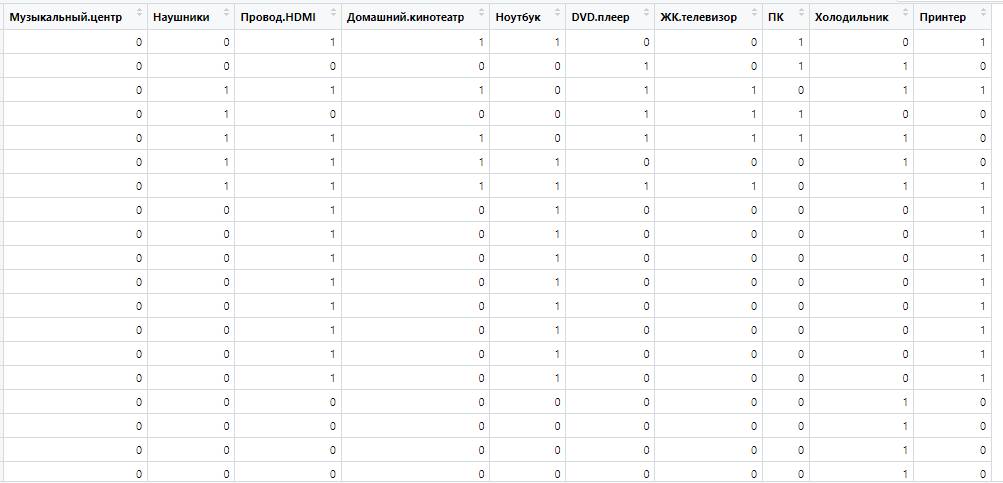
Далее нужно загрузить данные из csv файла



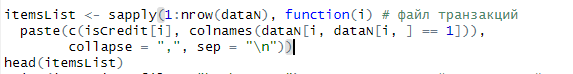
Поиск ассоциативных правил будет проводиться для записей с использованием кредита, поэтому необходимо отделить столбец «Кредит».



Ниже показана таблица без столбца «Кредит».



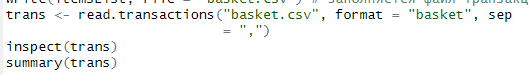
Для проведения ассоциативного анализа исходные данные преобразуются в файл транзакций.



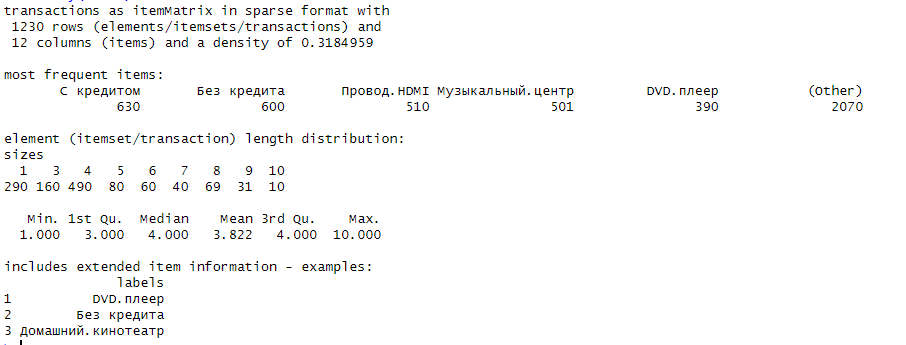
Далее идет заполнение файла транзакций.



Теперь нужно считать записанные ранее транзакции из файла.



С помощью команды summary(trans) можно посмотреть сводную информацию о наборе транзакций.



Транзакции: 1230 строк — каждая строка представляет собой отдельную транзакцию (набор товаров).

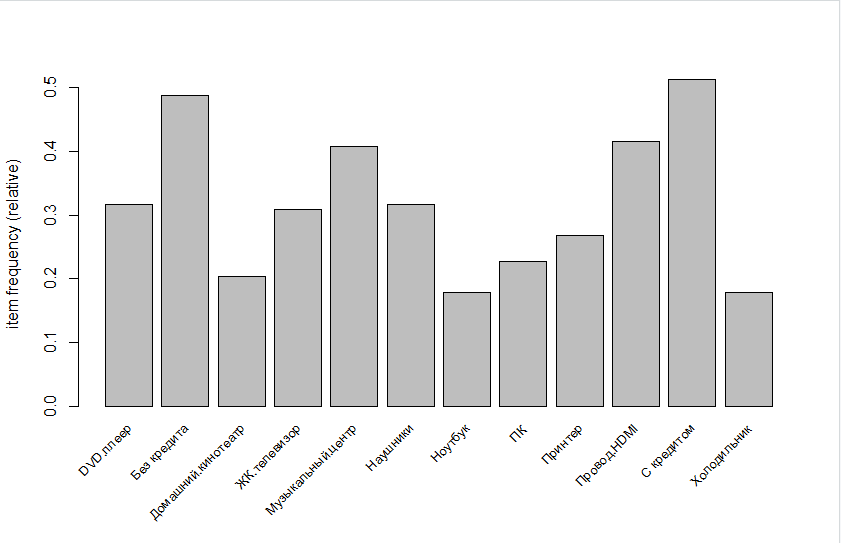
Товары: 12 уникальных позиций.

Плотность: 0.318 (31.8% элементов матрицы заполнены). Плотность показывает, какая часть ячеек в матрице содержит ненулевые значения Распределение длины транзакций

* Одноэлементные транзакции: 290
* Транзакции из 4 товаров: 490 (наиболее частая длина).
* Максимальная длина транзакции: 10 товаров.
* Средняя длина транзакции: 3.822. Большинство транзакций содержат от 3 до 4 товаров.

**Построение частотной диаграммы транзакций.**

****

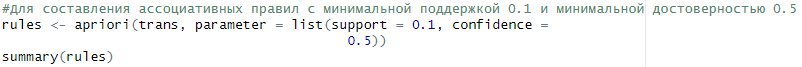


Частотная диаграмма транзакций показывает, насколько часто товары встречаются в наборе транзакций. По вертикальной оси отображается относительная частота появления товаров, а по горизонтальной — названия товаров.

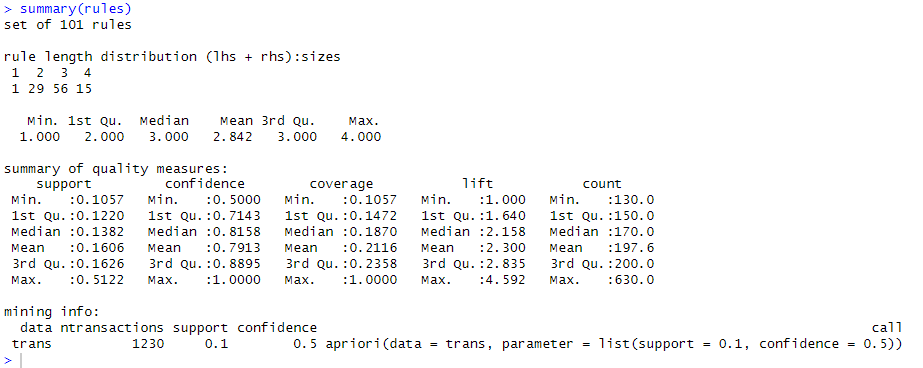
С кредитом и без кредита — встречаются почти в половине транзакций (около 0.5). Провод HDMI и Музыкальный центр также имеют высокую частоту.

Высокая частота категорий "С кредитом" и "Без кредита" указывает на значимость этих условий в транзакциях, кредиты являются важным фактором при покупке товаров.

**Cоставления ассоциативных правил с минимальной поддержкой 0.1 и минимальной достоверностью 0.5**

****

Cводной информации о полученных ассоциативных:



Длина правила — это сумма элементов в левой части (lhs) и правой части (rhs) правила. LHS (antecedent) — это условие или набор элементов (товаров), которые уже присутствуют в транзакции.

RHS (consequent) — это элемент (или элементы), которые "ассоциируются" с левой частью. Если в транзакции есть товары из LHS, то, с высокой вероятностью, будет и RHS.

Пример "Ноутбук", "С кредитом"→"Провод-HDMI" (Если покупатель приобрёл ноутбук в кредит, то есть высокая вероятность, что он купит и провод HDMI.):

* 1 правило имеет длину 1 (только один элемент, вероятно, только в rhs или lhs).
* 29 правил имеют длину 2 (один элемент слева и один справа).
* 56 правил имеют длину 3 (два элемента слева и один справа или наоборот).
* 15 правил имеют длину 4 (например, три элемента слева и один справа).

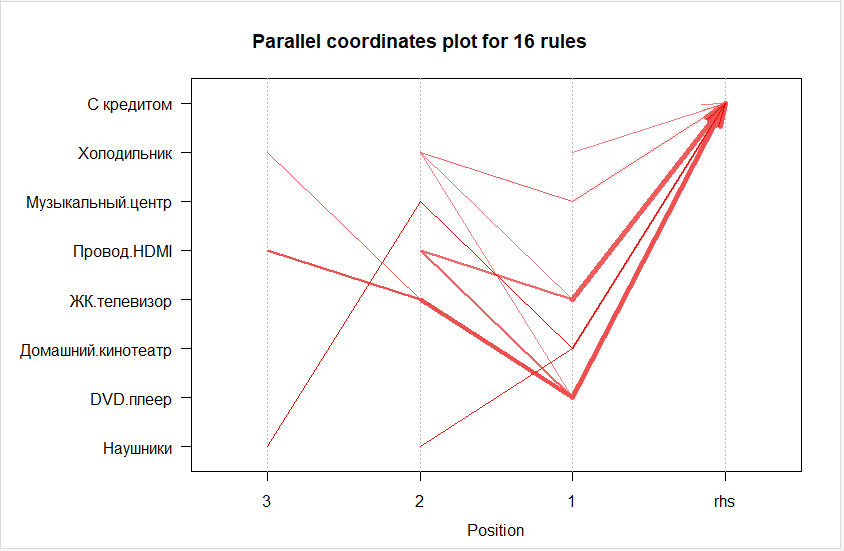
Медиана длины равна 3, что указывает на то, что большинство правил включает 2–3 элемента в левой части и один элемент в правой.

Средняя достоверность: 0.7913 (около 79% вероятности того, что правило верно, если левая часть выполняется). Высокая достоверность означает, что правила достаточно надёжные и показывают сильную связь между товарами.

Среднее значение Lift: 2.3, что говорит о том, что в среднем товары связаны друг с другом более чем в 2 раза чаще, чем случайно.

**Построение сети ассоциативных правил**

****

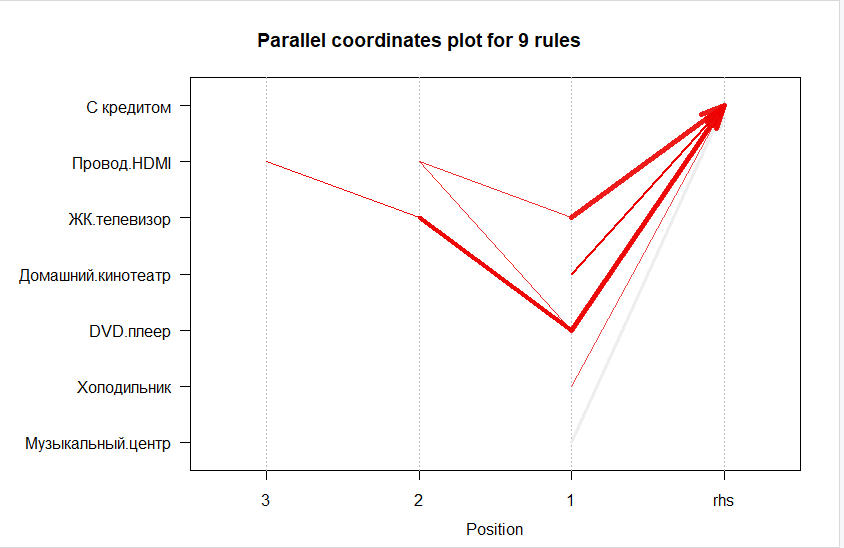
****

Ось Position (1, 2, 3) указывает позиции элементов в левой части (LHS) правила, а ось rhs показывает элемент, находящийся в правой части правила ( "С кредитом").

Линии разной толщины указывают на разный уровень поддержки (support) и достоверности (confidence). Более толстые линии соответствуют правилам с более высокими значениями поддержки или достоверности.

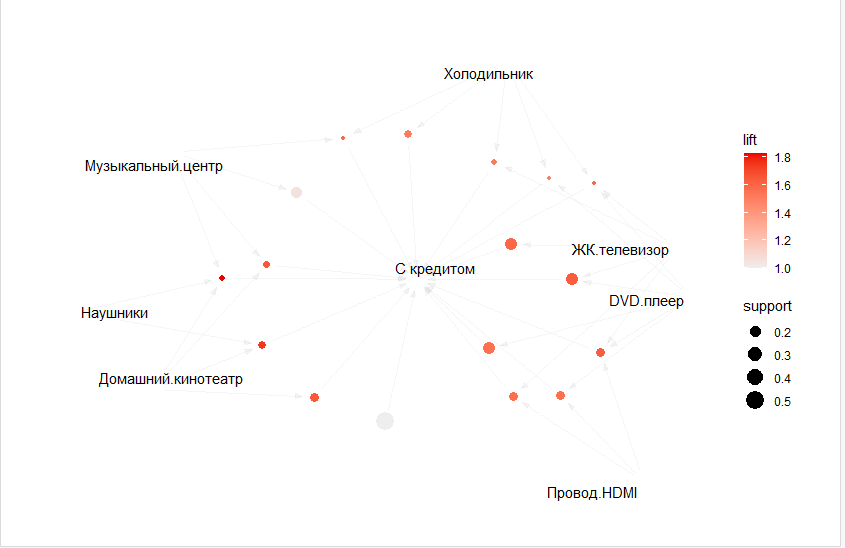
Каждая линия начинается на одной из позиций слева (1, 2 или 3), что показывает, какие товары (наушники, холодильник и т.д.) связаны с конечным результатом — "С кредитом".

****

****

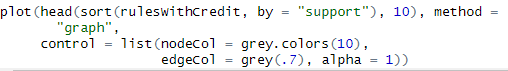
**Представления ассоциативных правил – граф.**

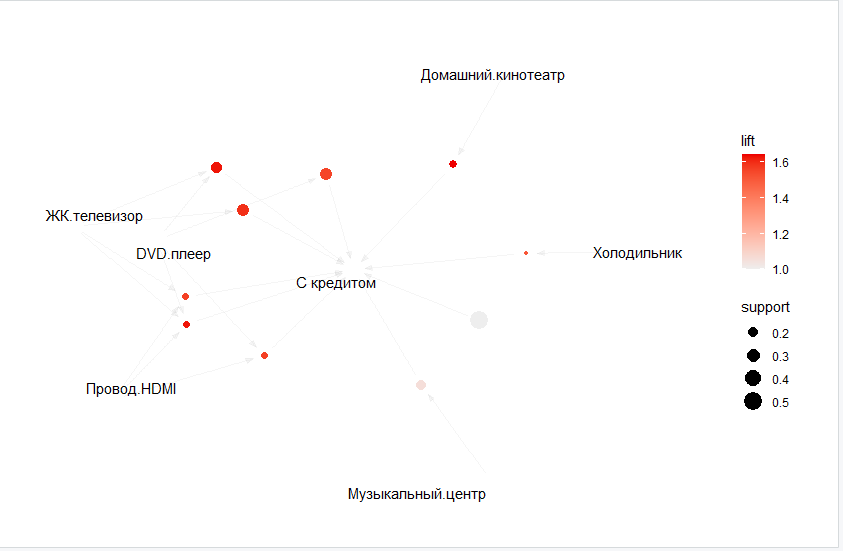
****

****

Размер узла указывает на поддержку (support) правила. Чем больше узел, тем чаще это правило встречается в данных. Lift > 1 означает, что вероятность покупки товара вместе с "С кредитом" выше, чем случайная вероятность.

Фильтрация первых 10 правил





**Заключение**

В ходе данной работы был проведён ассоциативный анализ, направленный на выявление скрытых закономерностей и связей между товарами, приобретаемыми в кредит. Основной целью было понять, какие товары чаще всего покупаются вместе и какие из них наиболее сильно связаны с транзакциями, осуществляемыми с использованием кредита.

Анализ показал, что такие дорогостоящие товары, как ЖК-телевизоры, домашние кинотеатры и холодильники, имеют наиболее сильную ассоциацию с покупками в кредит.

**Приложение**

library("arules")

library("arulesViz")

data <- read.delim(file = "lab3\_10.csv", sep = ",", header

= TRUE, row.names = 1)

View(data)

str(data)

dataN <- data[, -11] # Берем данные, исключая столбец Кредит

isCredit <- data$Кредит # Столбец "Кредит"

isCredit[isCredit == 1] <- "С кредитом"

isCredit[isCredit == 0] <- "Без кредита"

View(dataN)

itemsList <- sapply(1:nrow(dataN), function(i) # файл транзакций

paste(c(isCredit[i], colnames(dataN[i, dataN[i, ] == 1])),

collapse = ",", sep = "\n"))

head(itemsList)

write(itemsList, file = "basket.csv") # заполняется файл транзакций.

trans <- read.transactions("basket.csv", format = "basket", sep

= ",")

inspect(trans)

summary(trans)

itemFrequencyPlot(trans, cex.names = 0.8) #Построение частотной диаграммы транзакций

#Для составления ассоциативных правил с минимальной поддержкой 0.1 и минимальной достоверностью 0.5 для заданных транзакций

rules <- apriori(trans, parameter = list(support = 0.1, confidence =

0.5))

summary(rules)

rulesWithCredit <- subset(rules, subset = rhs

%in% "С кредитом") # транзакции с использованием кредита

plot(rulesWithCredit, method = "paracoord")

plot(head(sort(rulesWithCredit, by = "support"), 10), method =

"paracoord") # ограничение количества отображаемых правил

plot(rulesWithCredit, method = "graph",

control = list(nodeCol = grey.colors(10),

edgeCol = grey(.7), alpha = 1))

plot(head(sort(rulesWithCredit, by = "support"), 10), method =

"graph",

control = list(nodeCol = grey.colors(10),

edgeCol = grey(.7), alpha = 1))