

## РАЗРАБОТКА РЕКОМЕНДАТЕЛЬНОЙ МОДЕЛИ ПОДДЕРЖКИ ПРИНЯТИЯ РЕШЕНИЯ ПРИ ВЫБОРЕ ПРОДУКТОВ ПОЛЬЗОВАТЕЛЕМ

Квятковская И. Ю.<sup>1</sup>, Во Тхи Хуен Чанг<sup>2</sup>,  
Чан Куок Тоан<sup>3</sup>

(Астраханский государственный технический  
университет, Астрахань)

*Рекомендательные системы используются для прогнозирования предпочтений пользователей в отношении определенного продукта или услуги, а также для рекомендации пользователю подходящих продуктов или услуг. Многие методы, используемые в интеллектуальном анализе данных, связанные с классификацией или построением ассоциативных правил, применяются в рекомендательных системах. Предлагается новая рекомендательная модель, сочетающая ассоциативные правила и меры индекса статистической импликации. В предлагаемой модели меры поддержки и достоверности используются для создания ассоциативных правил, а мера индекса статистической импликации используется для фильтрации набора правил и ранжирования рекомендаций. Предложенные модель и алгоритмы использованы для построения рекомендательного результата по известному набору данных.*

Ключевые слова: рекомендательная система, рекомендательная модель, анализ статистической импликации, индекс статистической импликации, алгоритм Apriori, ассоциативные правила.

В настоящее время с быстрым развитием интернета и социальных сетей количество информации, к которой люди имеют доступ, увеличивается. Каждый день пользователи сталкиваются с многочисленными источниками информации: информацией, которой обмениваются по электронной почте, статьями в Интернете, сообщениями в социальных сетях, рекламной информацией с сайтов электронной коммерции [1–3, 5, 6, 13]. С увеличением всевозможной информации становится все труднее выбирать полезную информацию для принятия решений пользователями компьютеров и смарт-устройств. Рекомендательные системы, которые широко применяются во многих об-

---

<sup>1</sup> Ирина Юрьевна Квятковская, д.т.н., профессор (i.kvyatkovskaya@astu.org).

<sup>2</sup> Во Тхи Хуен Чанг (vthtrang@mail.ru).

<sup>3</sup> Чан Куок Тоан, к.т.н. (hoaiivan219@mail.ru).

ластях, рассматриваются как решение, помогающее пользователям эффективно выбирать информацию. Рекомендательная система способна автоматически анализировать информацию, классифицировать, выбирать и предоставлять пользователям интересующие продукты, товары и услуги посредством применения статистических и интеллектуальных методов (экспертные системы, нечеткие системы, системы поддержки принятия решений) [1–3]. Таким образом, исследование рекомендательных систем является актуальным.

Одной из основных функций рекомендательных систем является оценка рейтингов товаров, которые не были рассмотрены пользователями. Эта оценка обычно основывается на собственных или других отзывах пользователей. Для рекомендации будут использоваться продукты с наивысшими оценками.

Формализуем задачу рекомендации: пусть  $U = \{u_1, u_2, \dots, u_M\}$  – множество пользователей, а  $I = \{i_1, i_2, \dots, i_N\}$  – множество продуктов. Функция  $f(u_a, i_j)$  измеряет релевантность (или рейтинг) продукта  $i_j$  для пользователя  $u_a$ ,  $f: U \times I \rightarrow R$ , где  $R$  – упорядоченное множество. Для каждого пользователя  $u_a \in U$  необходимо найти такой продукт  $i_j \in I$ , чтобы функция  $f(u_a, i_j)$  имела максимальное значение:

$$(1) \quad \forall u_a \in U, i_j = \arg \max_{i_k \in I} f(u_a, i_k).$$

Задача рекомендации встроена в общую модель, как показано на рис. 1.

Среди рекомендательных моделей самыми успешными являются модели совместной фильтрации, такие как:

- рекомендательная модель совместной фильтрации на основе пользователей;
- рекомендательная модель совместной фильтрации на основе продукта;
- рекомендательная модель совместной фильтрации на основе ассоциативных правил.

		Продукты					
		$i_1$	$i_2$	$i_3$	$i_4$	...	$i_m$
Пользователь	$u_1$	?	4	4	1	...	?
	$u_2$	3	?	?	?	...	?
	$u_3$	4	?	?	3	...	3
	$u_4$	3	?	3	2	...	4
	...	...	...	...	...	...	...
	$u_n$	2	?	?	?	...	1
	$u_a$	?	?	4	2	...	3

Рис. 1. Общая схема рекомендации

Все модели учитывают, что взаимное влияние пары пользователей симметрично. На практике роли и взаимодействия между двумя пользователями часто асимметричны, что может создать определенную предвзятость в рекомендациях. Так более опытный пользователь-эксперт имеет большее влияние на других, менее опытный эксперт или новичок не сможет достичь противоположного эффекта. Существующие рекомендательные системы сосредоточены только на логическом решении вопроса о наличии или отсутствии взаимосвязи между пользователем и продуктами, однако не используют импликативное отношение «если А, то В» между ними, определяя степень соответствия на основе статистических данных о взаимозависимостях между пользователями и выбранными продуктами. В данном случае импликативное отношение рассматривается как отношение, основанное на знаниях о связях между пользователем и элементами данных, необходимых для выработки рекомендации. Использование асимметричного подхода представляет большой интерес для минимизации систематической ошибки из-за указанного выше различия в результатах рекомендаций.

Для учета асимметричного влияния пользователей и оценки импликативной связи между пользователем и продуктами используется метод анализа статистической импликации, или ме-

тод имплицативного статистического анализа (англ. «Statistical Implicative Analysis (SIA)», [8–10]).

Данный метод анализа данных позволяет обнаружить правила  $a \rightarrow b$  ( $a$  – атрибуты объектов, принадлежащих множеству  $A$ ,  $b$  – атрибуты объектов во множестве  $B$ ) в асимметричной форме «если  $a$ , то почти  $b$ » или «в какой степени  $b$  соответствует импликации  $a$ ». Целью метода является обнаружение тенденций в наборе атрибутов (переменных) с использованием меры индекса статистической импликации, формула которого приведена ниже. В отличие от других методов анализа данных, метод анализа статистической импликации ориентирован на асимметричные соотношения между переменными (когда привлекательная ценность правила  $a \rightarrow b$  отличается от привлекательной ценности правила  $b \rightarrow a$ ). Мера импликации используется для обнаружения правил, которые имеют сильную взаимосвязь импликации между атрибутами левой стороны и атрибутами правой части. Анализ статистической импликации применяется во многих областях, таких как образование, психология, информационные технологии и т. д.

Предположим, что набор  $E$  имеет  $n$  объектов или продуктов. Пусть  $A \subset E$  – подмножество объектов, удовлетворяющих атрибутам  $a$ ;  $B \subset E$  – подмножество объектов, удовлетворяющих атрибутам  $b$ ;  $\bar{A}$  (соответственно  $\bar{B}$ ) является дополнением  $A$  (соответственно  $B$ );  $n_A = |A|$ ,  $n_B = |B|$  – количество элементов множеств  $A$  и  $B$ ; число контрпримеров  $n_{A\bar{B}} = |A \cap \bar{B}|$  – это число объектов, которые удовлетворяют свойству  $a$ , но не удовлетворяют свойству  $b$ . Пусть  $X$  и  $Y$  – два случайных набора с количеством элементов  $n_X = n_A$  и  $n_Y = n_B$  соответственно. Предположим, что в процессе выборки случайная величина  $|X \cap \bar{Y}|$  следует распределению Пуассона с параметром

$$(2) \quad \lambda = \frac{n_A \cdot n_{\bar{B}}}{n}.$$

Правило  $a \rightarrow b$  называется приемлемым для заданного порога  $\alpha$ , если

$$(3) \quad Pr[|(X \cap \bar{Y})| \leq |A \cap \bar{B}|] \leq \alpha,$$

где  $Pr$  – вероятность случайной величины.

Рассмотрим случай  $n_B \neq 0$ . Тогда согласно распределению Пуассона вероятность случайной величины  $|X \cap \bar{Y}|$  определяется следующим образом:

$$(4) \quad Q(A, \bar{B}) = \left( |X \cap \bar{Y}| - \frac{n_A(n - n_B)}{n} \right) / \sqrt{\frac{n_A(n - n_B)}{n}}.$$

В эксперименте наблюдаемое значение  $q(A, \bar{B})$  для  $Q(A, \bar{B})$  определяется следующим образом:

$$(5) \quad q(A, \bar{B}) = \left( n_{A\bar{B}} - \frac{n_A(n - n_B)}{n} \right) / \sqrt{\frac{n_A(n - n_B)}{n}}.$$

Это значение называется индексом статистической импликации.

С использованием метода анализа статистической импликации разработана рекомендательная модель, основанная на индексе статистической импликации (ИСИ). Модель рассматривает взаимосвязь между набором условных атрибутов и набором атрибутов решения на основе асимметричного подхода. Модель выбирает ассоциативные правила, которые отфильтрованы по величине ИСИ, чтобы рекомендовать пользователю продукты. Рекомендательная модель определяется следующим образом:

$$(6) \quad RS = \{U, I, Ru, F\},$$

где  $U = \{u_1, u_2, \dots, u_M\}$  – множество из  $M$  пользователей;  $I = \{i_1, i_2, \dots, i_k, i_{k+1}, \dots, i_N\}$  – набор из  $N$  продуктов (атрибутов) для каждого пользователя, где  $I_{co} = \{i_1, i_2, \dots, i_k\}$  – набор условных продуктов (атрибутов), определяющих условия для выработки рекомендации;  $I_{re} = \{i_{k+1}, \dots, i_N\}$  – набор продуктов (атрибутов), определяющих решения;  $Ru = \{r_1, r_2, \dots, r_R\}$  – набор ассоциативных правил, выбранных для модели;  $F : U \times I \rightarrow Ru$  – вычислительные функции для поиска ассоциативных правил  $Ru_a = \{r_1, r_2, \dots, r_{RA}\}$ , где  $RA$  – число ассоциативных правил с рекомендательной ценностью для пользователя  $u_a$ , вычисленной на основе меры индекса статистической импликации.

Процесс работы рекомендательной модели, основанной на ИСИ, представлен на рис. 2. В основе ее лежит анализ транзакций – информации, хранящейся в базах данных продавцов товаров, о выборе (покупке) пользователем  $u \in U$  товара  $i \in I$ .



Рис. 2. Процесс работы рекомендательной модели на основе меры индекса статистической импликации

На основе предлагаемой рекомендательной модели разработана процедура для поддержки принятия решения по выбору продуктов для рекомендации пользователю.

*Входные данные:* набор данных транзакции (извлекаются из базы данных о пользователях и продуктах); набор значений условных атрибутов ( $I_{co}$ ) пользователя  $u_a$ ; набор продуктов (атрибутов), определяющих решения  $I_{re} = \{i_{k+1}, \dots, i_N\}$ .

*Выходные данные:* набор ассоциативных правил, поддерживающих выбор значения атрибута решения.

В процедуре выделены 4 главных этапа.

Этап 1. Создать множество правил ассоциации  $Ru$  на основе атрибута решения из набора данных транзакции.

Этап 2. Произвести определение значений параметров статистической импликации для множества правил ассоциации.  $n$ ;  $n_A$ ;  $n_B$ ;  $n_{A\bar{B}}$ .

Этап 3. Произвести расчет значения ИСИ для каждого ассоциативного правила  $r_j$ , принадлежащего множеству правил  $Ru$ .

Этап 4. Выбрать набор рекомендательных правил для пользователя  $u_a$ .

4.1. Для каждого правила ассоциации на основе атрибута решения правила  $r_j$ , принадлежащего набору правил  $Ru$ , выполнить: если ( $\langle$ Значение атрибутов левой части правила  $r_j \rangle \subseteq \subseteq \langle$ Множеству значений  $I_c \rangle$ ), то поместить правило ассоциации  $r_j$  в набор рекомендательных правил для пользователя  $u_a$ .

4.2. Сортировать набор рекомендательных правил для пользователя  $u_a$  в соответствии с ИСИ.

4.3. Выбрать правила с наивысшим значением ИСИ для рекомендации пользователю  $u_a$ .

*Конец.*

Для реализации этапа 1 разработан алгоритм генерации ассоциативных правил на основе атрибутов решения, основанный на алгоритме генерации ассоциативных правил Apriori [4] и примененный для правосторонних ассоциативных правил, содержащих атрибуты решения.

Сначала с использованием порога меры поддержки (англ. Support) ( $min\_sup$ ) для поиска часто встречающихся наборов атрибутов осуществляется поиск набора из одного атрибута ( $L_1$ ). Далее  $L_1$  используется для нахождения набора из двух атрибутов ( $L_2$ ) и так до тех пор, пока набор из  $p$  атрибутов ( $L_p$ ) больше не будет найден.

Далее на основе порога меры достоверности (англ. Confidence) ( $min\_conf$ ) формируются правила ассоциации рекомендательной модели из часто встречающихся атрибутов.

*Псевдокод алгоритма генерации правила ассоциации на основе атрибутов решения.*

*Входные данные:* транзакции о пользователях и продуктах)  $T$ ,  $I_c$  – набор условных продуктов (атрибутов), определяющих условия для выработки рекомендации;  $I_r$  – набор продуктов (атрибутов), определяющих решения;  $min\_sup$ ,  $min\_conf$  – пороговые значения показателей «поддержка» и «достоверность», характеризующих ассоциативные правила.

*Выходные данные:* набор ассоциативных правил для рекомендательной модели.

*Начало*

Шаг 1. Просмотреть весь набор транзакций  $T$ , чтобы определить поддержку (меру Support)  $S$  (процент транзакций, содержащих определенный набор данных) для набора из одного продукта:

$$(7) S_a = \frac{\text{Количество транзакций, содержащих } a}{|T|},$$

где  $a$  – заданный продукт,  $|T|$  – общее количество транзакций, равное количеству пользователей. Если в базе данных по каждому пользователю хранится единственная транзакция, то  $|T| = M$ .

Далее, сравнивая  $S_a$  с минимальным пороговым значением  $min\_sup$ , отобрать подмножество наборов из одного продукта ( $L_1$ ).

Шаг 2. Далее, используя набор  $L_{p-1}$ , полученный на предыдущем шаге, сгенерировать  $L_p$  – подмножество наборов из  $p$  продуктов, используя для отбора заданное минимальное пороговое значение  $min\_sup$ .

Шаг 3. Повтор действия с шага 2 до тех пор, пока сгенерированный набор продуктов не будет пуст.

Окончательно получим  $L = \{L_1, L_2, \dots, L_p\}$  – множество наборов продуктов.

Шаг 4. Разбить каждый набор  $L_t \in L$  ( $t \in \{1, \dots, p\}$ ) на элементарные одноэлементные непустые подмножества  $V \subset L_t$ .

Шаг 5. Для каждого непустого подмножества  $V \subset L_t$  сгенерировать правила  $Ru_r = \{V \rightarrow (L_t \setminus V) \mid (L_t \setminus V) \in I_r\}$ . Вычислить  $Conf(V \rightarrow (L_t \setminus V)) \geq min\_conf$ , где  $Conf(V \rightarrow (L_t \setminus V))$  – достоверность ассоциативного правила  $V \rightarrow (L_t \setminus V)$ , определяемая по формуле:

$$(8) Conf(V \rightarrow (L_t \setminus V)) = \frac{S(V)}{S(L_t)},$$

где  $S(L_t)$  – поддержка набора  $L_t$ ;  $S(V)$  – поддержка набора  $V$ .

*Конец.*

**Пример 1.** Используем набор из трех продуктов  $I = \{i_1, i_2, i_3\}$ . Зададим  $I_c = \{i_1, i_2\}$  – набор условных продуктов (атрибутов), определяющих условия для выработки рекомендации,  $I_r = \{i_3\}$  – набор продуктов (атрибутов), определяющих ре-

шения. Например,  $i_1 = \{1; 2; 3\}$ ;  $i_2 = \{0,5; 1,0; 1,5\}$ ,  $i_3 = \{A; B\}$  (таблица 1).

Таблица 1. Набор исходных данных

№ транзакции	Продукты		
	$i_1$	$i_2$	$i_3$
1	2	1,0	A
2	1	1,5	A
3	3	0,5	B
4	3	0,5	B
5	2	1,0	A
6	1	1,5	A
7	1	1,5	A
8	2	1,0	A
9	3	0,5	B
10	2	1,0	A

Применение алгоритма генерации правила ассоциации на основе атрибута решения с заданными  $min\_sup = 0,01$  и  $min\_conf = 0,5$  включает следующие шаги:

Найти среди транзакций часто встречающиеся наборы из одного элемента, вычислить поддержку  $S$  для каждого набора. Результат вычисления представлен в таблице 2.

Таблица 2. Результат вычисления поддержки для набора из одного элемента

Набор из 1 элемента	Поддержка $S$
$\{i_1 = 1\}$	0,3
$\{i_1 = 2\}$	0,4
$\{i_1 = 3\}$	0,3
$\{i_2 = 1,0\}$	0,4
$\{i_2 = 1,5\}$	0,3
$\{i_2 = 0,5\}$	0,3
$\{i_3 = A\}$	0,7
$\{i_3 = B\}$	0,3

Сравнить значение поддержки с заданным  $min\_sup = 0,01$ . Результат формирования набора  $L_1$  представлен в таблице 3.

Таблица 3. Формирование  $L_1$

Набор из 1 элемента	Поддержка
{i1 = 1}	0,3
{i1 = 2}	0,4
{i1 = 3}	0,3
{i2 = 1,0}	0,4
{i2 = 1,5}	0,3
{i2 = 0,5}	0,3
{i3 = A}	0,7
{i3 = B}	0,3

Найти часто встречающиеся наборы из двух элементов. Вычислить поддержку  $S$  для 2-элементного набора. Результаты вычисления представлены в таблице 4.

Таблица 4. Результаты вычисления поддержки для набора из двух элементов

Набор из 2 элементов	Поддержка $S$
{i1 = 1; i1 = 2}	0,0
{i1 = 1; i1 = 3}	0,0
{i1 = 1; i2 = 1,0}	0,0
{i1 = 1; i2 = 1,5}	0,3
{i1 = 1; i2 = 0,5}	0,0
{i1 = 1; i3 = A}	0,3
{i1 = 1; i3 = B}	0,0
{i1 = 2; i1 = 3}	0,0
{i1 = 2; i2 = 1,0}	0,4
{i1 = 2; i2 = 1,5}	0,0
{i1 = 2; i2 = 0,5}	0,0
{i1 = 2; i3 = A}	0,4
{i1 = 2; i3 = B}	0,0
{i1 = 3; i2 = 1,0}	0,0
{i1 = 3; i2 = 1,5}	0,0
{i1 = 3; i2 = 0,5}	0,3
{i1 = 3; i3 = A}	0,0
{i1 = 3; i3 = B}	0,3
{i2 = 1,0; i2 = 1,5}	0,0
{i2 = 1,0; i2 = 0,5}	0,0
{i2 = 1,0; i3 = A}	0,4

Таблица 4 (продолжение)

{i2 = 1,0; i3 = B}	0,0
{i2 = 1,5; i2 = 0,5}	0,0
{i2 = 1,5; i3 = A}	0,3
{i2 = 1,5; i3 = B}	0,0
{i2 = 0,5; i3 = A}	0,0
{i2 = 0,5; i3 = B}	0,3
{i3 = A; i3 = B }	0,0

Выбрать наборы, сравнивая значение поддержки с заданным  $min\_sup = 0,01$ . Результат формирования  $L_2$  представлен в таблице 5.

Таблица 5 Формирование  $L_2$

Набор из 2 элементов	Поддержка
{i1 = 1; i2 = 1,5}	0,3
{i1 = 1; i3 = A}	0,3
{i1 = 2; i2 = 1,0}	0,4
{i1 = 2; i3 = A}	0,4
{i1 = 3; i2 = 0,5}	0,3
{i1 = 3; i3 = B}	0,3
{i2 = 1,0; i3 = A}	0,4
{i2 = 1,5; i3 = A}	0,3
{i2 = 0,5; i3 = B}	0,3

Найти часто встречающиеся наборы из трёх элементов. Вычислить поддержку  $S$  для 3-элементного набора. Результаты вычисления представлены в таблице 6.

Таблица 6. Результаты создания частого набора из трёх элементов

Набор из 3 элементов	Поддержка $S$
{i1 = 1; i2 = 1,5; i3 = A}	0,3
{i1 = 1; i2 = 1,5; i3 = B}	0,0
{i1 = 1; i2 = 1,0; i3 = A}	0,0
{i1 = 1; i2 = 1,0; i3 = B}	0,0
{i1 = 1; i2 = 0,5; i3 = A}	0,0
{i1 = 1; i2 = 0,5; i3 = B}	0,0

Таблица 6 (продолжение)

{i1 = 2; i2 = 1,0; i3 = A}	0,3
{i1 = 2; i2 = 1,0; i3 = B}	0,0
{i1 = 2; i2 = 1,5; i3 = A}	0,0
{i1 = 2; i2 = 1,5; i3 = B}	0,0
{i1 = 2; i2 = 0,5; i3 = A}	0,0
{i1 = 2; i2 = 0,5; i3 = B}	0,0
{i1 = 3; i2 = 0,5; i3 = A}	0,0
{i1 = 3; i2 = 0,5; i3 = B}	0,3
{i1 = 3; i2 = 1,5; i3 = A}	0,0
{i1 = 3; i2 = 1,5; i3 = B}	0,0
{i1 = 3; i2 = 1,0; i3 = A}	0,0
{i1 = 3; i2 = 1,0; i3 = B}	0,0

Сравнить значение поддержки с заданным  $min\_sup = 0,01$ . Результат формирования  $L_3$  представлен в таблице 7.

Таблица 7 Формирование  $L_3$

Набор из 3 элементов	Поддержка
{i1 = 1; i2 = 1,5; i3 = A}	0,3
{i1 = 2; i2 = 1,0; i3 = A}	0,3
{i1 = 3; i2 = 0,5; i3 = B}	0,3

Окончательно был сгенерирован набор ассоциативных правил на основе атрибутов, определяющих решения, включая 9 правил, представленных в таблице 8.

Таблица 8. Результат создания ассоциативных правил на основе атрибута решения

№	Ассоциативные правила	Поддержка	Достоверность
1	{i1 = 1} => {i3 = A}	0,3	1,0
2	{i2 = 1,5} => {i3 = A}	0,3	1,0
3	{i1 = 3} => {i3 = B}	0,3	1,0
4	{i2 = 0,5} => {i3 = B}	0,3	1,0
5	{i2 = 1} => {i3 = A}	0,4	1,0
6	{i1 = 2} => {i3 = A}	0,4	1,0
7	{i1 = 1; i2 = 1,5} => {i3 = A}	0,3	1,0
8	{i1 = 3; i2 = 0,5} => {i3 = B}	0,3	1,0
9	{i1 = 2; i2 = 1} => {i3 = A}	0,4	1,0

Для реализации шага 2 процедуры разработана методика формирования меры интенсивности статистической импликации для правила ассоциации, состоящая из следующих шагов:

Шаг 1. Производится преобразование набора данных транзакции в разреженную бинарную матрицу.

Шаг 2. Производится преобразование набора правил ассоциации в разреженную бинарную матрицу.

Шаг 3. Производится вычисления значений параметров статистической импликации правила ассоциации.

Шаг 4. Производится вычисления значения индекса статистической импликации для правила ассоциации.

Для определения значений параметров  $n$ ;  $n_A$ ;  $n_B$ ;  $n_{A\bar{B}}$  каждого правила ассоциации предлагается метод преобразования набора данных транзакции и правил ассоциации в разреженную бинарную матрицу. Это преобразование преследует две основные цели.

Во-первых, большинство алгоритмов генерации ассоциативных правил основано на бинарных данных, что удобно при экспериментальной реализации.

Во-вторых, с разреженной бинарной структурой матрицы алгоритм может решить проблему пространства для хранения больших наборов данных и увеличить скорость обработки алгоритмов модели за счет того, что алгоритмы обрабатывают данные только во внутренней памяти, вместо того чтобы обращаться к внешней.

*Комментарии:*

*По шагу 1 методики.* Преобразование набора данных транзакции в разреженную бинарную матрицу  $BM$ , состоящую из  $n$  строк и  $m$  столбцов, где  $n$  – общее количество транзакций в наборе данных,  $m$  – сумма значений всех атрибутов, присутствующих в наборе данных транзакции. Каждая транзакция преобразуется в строку матрицы, каждое значение атрибутов в наборе данных транзакции становится столбцом матрицы. Значения матрицы  $BM$  определяются по следующему принципу: если транзакция в строке  $i$  содержит соответствующее значение атрибута в столбце  $j$ ,  $BM_{i,j} = 1$ , в противном случае  $BM_{i,j} = 0$ .

**Пример 2.** Из набора данных транзакции, представленного в примере 1, преобразуем данные в бинарную матрицу  $BM(10 \times 8)$ . Поскольку данные транзакции в строке 1 содержат только значения  $i_1 = 2$ ;  $i_2 = 1,0$ ;  $i_3 = A$ ,  $BM_{1,2} = 1$ ,  $BM_{1,5} = 1$ ,  $BM_{1,7} = 1$ , а остальные элементы строки 1 равны 0. Аналогично определены значения для остальных строк матрицы. Результаты представлены в таблице 9.

Таблица 9. Результаты преобразования данных транзакции в бинарную матрицу

$i_1=1$	$i_1=2$	$i_1=3$	$i_2=1,5$	$i_2=1,0$	$i_2=0,5$	$i_3=A$	$i_3=B$
0	1	0	0	1	0	1	0
1	0	0	1	0	0	1	0
0	0	1	0	0	1	0	1
0	0	1	0	0	1	0	1
0	1	0	0	1	0	1	0
1	0	0	1	0	0	1	0
1	0	0	1	0	0	1	0
0	1	0	0	1	0	1	0
0	0	1	0	0	1	0	1
0	1	0	0	1	0	1	0

По шагу 2 методики. Преобразование набора правил ассоциации в разреженную бинарную матрицу. Данные, преобразованные в форму бинарной матрицы набора правил, делятся на три части: полную матрицу бинарных правил, левую бинарную матрицу правил и правую бинарную матрицу правил. В каждой бинарной матрице преобразование выполняется в соответствии со следующим принципом: каждое правило ассоциации преобразуется в строку матрицы, каждый атрибут, появляющийся в левой или правой части правила ассоциации, преобразуется в столбцы в матрице. Значения матрицы определяются аналогично.

**Пример 3.** Из набора правил ассоциации, представленного в примере 1, преобразуем данные в бинарную матрицу следующим образом:

Полная бинарная матрица правил ассоциации  $BMF$  имеет 9 строк и 8 столбцов. Поскольку правило ассоциации в строке 1

содержит только значения атрибутов  $i_1 = 1$ ,  $i_3 = A$ , поэтому  $BMF_{1,1} = 1$ ,  $BMF_{1,7} = 1$ . Аналогично определяются значения для остальных строк матрицы. Результаты представлены в таблице 10.

Таблица 10. Результат преобразования правила ассоциации в бинарную матрицу

$i_1=1$	$i_1=2$	$i_1=3$	$i_2=1,5$	$i_2=1,0$	$i_2=0,5$	$i_3=A$	$i_3=B$
1	0	0	0	0	0	1	0
0	0	0	1	0	0	1	0
0	0	1	0	0	0	0	1
0	0	0	0	0	1	0	1
0	0	0	0	1	0	1	0
0	1	0	0	0	0	1	0
1	0	0	1	0	0	1	0
0	0	1	0	0	1	0	1
0	1	0	0	1	0	1	0

Левая бинарная матрица правил ассоциации  $BML$  имеет размерность  $(9 \times 6)$ . В первой строке  $BML_{1,1} = 1$ , остальные элементы равны 0, поскольку левая часть правила ассоциации в строке 1 содержит только значение атрибута  $i_1 = 1$ . Аналогично определяем значения для остальных строк матрицы. Результаты преобразования представлены в таблице 11.

Таблица 11. Результат преобразования левой части правила ассоциации в бинарную матрицу

$i_1=1$	$i_1=2$	$i_1=3$	$i_2=1,5$	$i_2=1,0$	$i_2=0,5$
1	0	0	0	0	0
0	0	0	1	0	0
0	0	1	0	0	0
0	0	0	0	0	1
0	0	0	0	1	0
0	1	0	0	0	0
1	0	0	1	0	0
0	0	1	0	0	1
0	1	0	0	1	0

Правая бинарная матрица правил ассоциации  $BMR$  имеет размерность  $(9 \times 2)$ . Поскольку правая часть правила ассоциации в строке 1 содержит только значение атрибута  $i_3 = A$ , то  $BML_{1,1} = 1$ , а остальные элементы строки 1 равны 0. Аналогично определяем значения для остальных строк матрицы. Результаты преобразования представлены в таблице 12.

Таблица 12. Результат преобразования правой части правила ассоциации в бинарную матрицу

$i_3=A$	$i_3=B$
1	0
1	0
0	1
0	1
1	0
1	0
1	0
0	1
1	0

По шагу 3 методики. Определение значений параметров статистической импликации каждого правила ассоциации. После вычисления бинарной матрицы данных, бинарных матриц правил ассоциации, следующим шагом будет определение значений параметров  $n$ ;  $n_A$ ;  $n_B$ ;  $n_{AB}$  для каждого правила в выбранном наборе правил в соответствии со следующим принципом:

– параметр  $n$ :  $n$  – количество строк бинарной матрицы данных  $BM$ ;

– параметр  $n_A$ : для каждой строки левой бинарной матрицы правил  $BML$  идентифицируются столбцы со значением 1. В соответствии с этими столбцами в полной бинарной матрице в каждом столбце определяют общее количество строк со значением 1, которое присваивают  $n_A$ ;

– параметр  $n_B$ : для каждой строки правой бинарной матрицы правил  $BMR$  идентифицируются столбцы со значением 1. В соответствии с этими столбцами в полной бинарной матрице в каждом столбце определяют общее количество строк со значением 1 и присваивают его  $n_B$ ;

– параметр  $n_{A\bar{B}}$ : для каждой строки полной бинарной матрицы правил *ВМФ* идентифицируются столбцы со значением 1. В соответствии с этими столбцами в полной бинарной матрице в каждом столбце определяют общее количество строк со значением 1 и присваивают его к  $n_{A\bar{B}}$ . Тогда,  $n_{A\bar{B}} = n_A - n_{AB}$ .

**Пример 4.** Из результата преобразования в бинарную матрицу в примере 2 и примере 3 переходим к определению значений параметров статистической импликации для первого правила следующим образом:

$$n = 10; n_A = 3; n_B = 7; n_{A\bar{B}} = 0.$$

Аналогично для остальных правил определяем значения параметров  $n$ ;  $n_A$ ;  $n_B$ ;  $n_{A\bar{B}}$ , представленные в таблице 13.

Таблица 13. Значения параметров  $n$ ;  $n_A$ ;  $n_B$ ;  $n_{A\bar{B}}$  для каждого ассоциативного правила

№	Ассоциативное правило	$n$	$n_A$	$n_B$	$n_{A\bar{B}}$
1	$\{i_1 = 1\} \Rightarrow \{i_3 = A\}$	10	3	7	0
2	$\{i_2 = 1,5\} \Rightarrow \{i_3 = A\}$	10	3	7	0
3	$\{i_1 = 3\} \Rightarrow \{i_3 = B\}$	10	3	3	0
4	$\{i_2 = 0,5\} \Rightarrow \{i_3 = B\}$	10	3	3	0
5	$\{i_2 = 1\} \Rightarrow \{i_3 = A\}$	10	4	7	0
6	$\{i_1 = 2\} \Rightarrow \{i_3 = A\}$	10	4	7	0
7	$\{i_1 = 1; i_2 = 1,5\} \Rightarrow \{i_3 = A\}$	10	3	7	0
8	$\{i_1 = 3; i_2 = 0,5\} \Rightarrow \{i_3 = B\}$	10	3	3	0
9	$\{i_1 = 2; i_2 = 1\} \Rightarrow \{i_3 = A\}$	10	4	7	0

По шагу 4 методики. Значения индекса статистической импликации для каждого ассоциативного правила, вычисленные по формуле (5), представлены в таблице 14.

Таблица 14. Значения индекса статистической импликации для каждого ассоциативного правила

№	Ассоциативное правило	$q(a,b)$
1	$\{i_1 = 1\} \Rightarrow \{i_3 = A\}$	-0,949
2	$\{i_2 = 1,5\} \Rightarrow \{i_3 = A\}$	-0,949
3	$\{i_1 = 3\} \Rightarrow \{i_3 = B\}$	-1,449
4	$\{i_2 = 0,5\} \Rightarrow \{i_3 = B\}$	-1,449
5	$\{i_2 = 1\} \Rightarrow \{i_3 = A\}$	-0,949
6	$\{i_1 = 2\} \Rightarrow \{i_3 = A\}$	-1,095
7	$\{i_1 = 1; i_2 = 1,5\} \Rightarrow \{i_3 = A\}$	-0,949
8	$\{i_1 = 3; i_2 = 0,5\} \Rightarrow \{i_3 = B\}$	-1,449
9	$\{i_1 = 2; i_2 = 1\} \Rightarrow \{i_3 = A\}$	-1,095

**Пример 5.** Для применимости модели проводился эксперимент на наборе данных Lenses [14], в котором хранятся атрибуты для выбора контактных линз. Он включает 24 записи для выбора контактных линз, 5 атрибутов (от  $i_1$  до  $i_4$  – атрибуты условия,  $i_5$  – атрибут решения). Здесь  $i_1$  (возраст пациента) включает три значения: 1 (молодой), 2 (в периоде перед дальнозоркостью), 3 (имеющий дальнозоркость);  $i_2$  (рецепт очков) включает два значения: 1 (близорукость), 2 (дальнозоркость);  $i_3$  (астигматизм) включает два значения: 1 (нет), 2 (да);  $i_4$  (скорость слезопroduкции) включает два значения: 1 (снижено), 2 (нормально). Далее,  $i_5$  – это атрибут решения, используемый для классификации трех значений: 1 (пациент должен носить жесткие контактные линзы), 2 (пациенту следует подобрать мягкие контактные линзы), 3 (пациенту не следует надевать контактные линзы). Подробное содержание набора данных представлено в таблице 15.

Производится создание ассоциативных правил для этого набора данных с использованием алгоритма генерации ассоциативных правил (алгоритм Apriori) на основе атрибутов решения (Поддержка = 0,15 и Достоверность = 0,8, исходя из рекомендаций [12]). Результаты, полученные с помощью набора из 10 правил рекомендательной системы, представлены в таблице 16.

Таблица 15. Подробное содержание набора данных *Lenses*

№	$i_1$	$i_2$	$i_3$	$i_4$	$i_5$
1	1	1	1	1	3
2	1	1	1	2	2
3	1	1	2	1	3
4	1	1	2	2	1
5	1	2	1	1	3
6	1	2	1	2	2
7	1	2	2	1	3
8	1	2	2	2	1
9	2	1	1	1	3
10	2	1	2	1	3
11	2	1	2	1	3
12	2	1	2	2	1
13	2	2	1	1	3
14	2	2	1	2	2
15	2	2	2	1	3
16	2	2	2	2	3
17	3	1	1	1	3
18	3	1	1	2	3
19	3	1	2	1	3
20	3	1	2	2	1
21	3	2	1	1	3
22	3	2	1	2	2
23	3	2	2	1	3
24	3	2	2	2	3

Таблица 16. Набор ассоциативных правил, сгенерированный для набора данных *Lenses*

№	Ассоциативные правила
1	$\{i_4 = 1\} \Rightarrow \{i_5 = 3\}$
2	$\{i_3 = 1; i_4 = 2\} \Rightarrow \{i_5 = 2\}$
3	$\{i_1 = 1; i_4 = 1\} \Rightarrow \{i_5 = 3\}$
4	$\{i_1 = 2; i_4 = 1\} \Rightarrow \{i_5 = 3\}$
5	$\{i_3 = 3; i_4 = 1\} \Rightarrow \{i_5 = 3\}$
6	$\{i_2 = 1; i_4 = 1\} \Rightarrow \{i_5 = 3\}$
7	$\{i_3 = 1; i_4 = 1\} \Rightarrow \{i_5 = 3\}$
8	$\{i_3 = 2; i_4 = 1\} \Rightarrow \{i_5 = 3\}$
9	$\{i_2 = 2; i_4 = 1\} \Rightarrow \{i_5 = 3\}$
10	$\{i_2 = 1; i_3 = 2\} \Rightarrow \{i_5 = 3\}$

Результат расчета значения меры ИСИ представлены в таблице 17.

Таблица 17. Значения меры ИСИ

№	ИСИ
1	-2,121
2	-1,721
3	-1,225
4	-1,225
5	-1,225
6	-1,500
7	-1,500
8	-1,500
9	-1,500
10	-0,833

На основе значения меры ИСИ модель дает рекомендательный результат. Например, при вводе значений условного признака  $\{i_1 = 1; i_2 = 2; i_3 = 2; i_4 = 1\}$  модель выдаст рекомендации, представленные в таблице 18.

Таблица 18. Результаты рекомендации при вводе значений условного признака  $\{i_1 = 1, i_2 = 2, i_3 = 2, i_4 = 1\}$

Ассоциативные правила	ИСИ
$\{i_4 = 1\} \Rightarrow \{i_5 = 3\}$	-2,121
$\{i_3 = 3; i_4 = 1\} \Rightarrow \{i_5 = 3\}$	-1,500
$\{i_1 = 2; i_4 = 1\} \Rightarrow \{i_5 = 3\}$	-1,500
$\{i_2 = 1; i_3 = 2\} \Rightarrow \{i_5 = 3\}$	-0,833

Выбираем ассоциативное правило с наибольшим значением ИСИ (-0,833), для которого  $i_5 = 3$ , следовательно, пациенту не следует использовать контактные линзы.

Оценка точности предложенной рекомендательной модели проведена на основе сравнения результатов рекомендаций модели с выбором пользователя [11]. Точность представляет собой оценку релевантности рекомендаций пользователю и вычисляется как отношение количества рекомендованных продуктов,

выбранных пользователем, к общему количеству рекомендованных продуктов.

Для оценки точности предлагаемой рекомендательной модели проводился эксперимент на наборе открытых данных MSWeb [7] о пользователях Microsoft, посетивших веб-сайты в течение одной недели, представленном в цифровом виде на сайте [www.microsoft.com](http://www.microsoft.com). Набор содержит данные о 38000 анонимных пользователей, получающих доступ к 294 исходным веб-адресам, и представлен в виде двоичной матрицы из 32711 строк, 294 столбцов и 98653 значений рейтинга.

Для оценки эффективности предложенной модели (PC\_ИСИ) в работе сравнивается точность предлагаемой модели с точностями рекомендательных моделей совместной фильтрации, таких как рекомендательная модель совместной фильтрации на основе пользователей (PC\_Пол) [11], рекомендательная модель совместной фильтрации на основе продукта (PC\_Прод) [11] и рекомендательная модель совместной фильтрации на основе ассоциативных правил (PC\_АП) [11]. Результаты оценки точности рекомендательных моделей для заданного набора данных представлены в таблице 19.

Таблица.19. Точности рекомендательных моделей

Рекомендательная модель	PC Пол	PC Прод	PC АП	PC ИСИ
Точность	0,425	0,401	0,501	0,699

Построена сравнительная диаграмма на основе результатов оценки, представленная на рис. 3.

По результатам оценки предлагаемая авторами рекомендательная модель имеет более высокую точность, чем остальные модели.

В работе разработана рекомендательная модель, основанная на асимметричном подходе с использованием ассоциативных правил, меры ИСИ. Преодолены недостатки традиционных рекомендательных моделей, учитывающих взаимное влияние пары пользователей как симметричное. Данная модель, в частности, использует асимметричную взаимосвязь между атрибутами условия и атрибутами решения в одном и том же пользовательском объекте. По значению меры ИСИ в выбранном наборе ас-

социативных правил модель дает рекомендательные результаты, чтобы помочь пользователям выбрать значения для атрибутов решения.

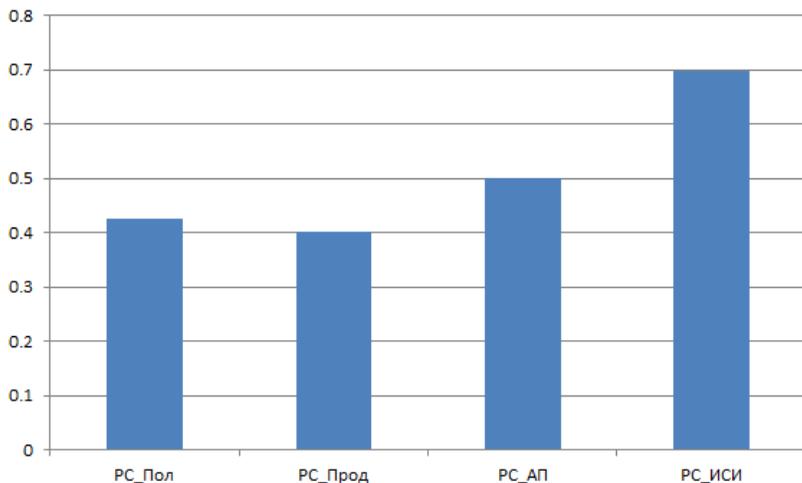


Рис.3 Сравнение точности рекомендательных моделей

## ЛИТЕРАТУРА

1. БРАХА Ш., ЛИОР Р. *Создание эффективных рекомендательных систем* // ISBN978-1-4419-0047-0 –2012 г. Архивировано из оригинала 01.05.2014.
2. БХАСКЕР Б.; СРИКУМАР К. *Рекомендательные системы в электронной коммерции* // ЧАШКА. – ISBN978-0-07-068067-8. Архивировано из оригинала 01.09.2010.
3. ФРАНЧЕСКО Р., ЛИОР Р., БРАХА Ш., КАНТОР П.Б. *Справочник рекомендательных систем.* – Springer. ISBN978-0-387-85819-7 – 2011.
4. ШИТИКОВ В.К., МАСТИЦКИЙ С.Э. *Классификация, регрессия, алгоритмы Data Mining с использованием R.* – 2017. – Электронная книга. – URL: <https://github.com/ranalytics/data-mining>.

5. ШУРШЕВ В.Ф., БУЙ Л.В. *Методика выбора сканирующих приемников и трансиверов по основным характеристикам* // Вестник Астраханского государственного технического университета. Серия: Управление, вычислительная техника и информатика. – 2013. – №2. – С. 45–51.
6. ШУРШЕВ В.Ф., КОЧКИН Г.А., КОЧКИНА В.Р. *Модель системы поддержки принятия решений на основе рассуждений по прецедентам* // Вестник Астраханского государственного технического университета. Серия: Управление, вычислительная техника и информатика. – 2013. – №2. – С. 175–183.
7. BREESE J.S., HECKERMAN D., KADIE C.M. *Anonymous web data from www.microsoft.com* // Microsoft Research, Redmond WA, 98052-6399, USA, 1998. – URL: <https://kdd.ics.uci.edu/databases/msweb/msweb.html>.
8. GRAS R. *Contribution à l'étude expérimentale et à l'analyse de certaines acquisitions cognitives et de certains objectifs en didactique des mathématiques*. – PhD thesis, Université de Rennes 1, 1979.
9. GRAS R., KUNTZ P. *An overview of the Statistical Implicative Analysis (SIA) development* // Statistical Implicative Analysis - Studies in Computational Intelligence, Springer-Verlag. – 2008. – P. 11–40.
10. GRAS R., KUNTZ P., GREFFARD N. *Notion of Implicative Fields in Statistical Implicative Analysis* // VIII Int. Conf. A.S.I. Analyse Statistique Implicative - Statistical Implicative Analysis Radès (Tunisie), Novembre, 2015. – P. 29–46.
11. HAHSLER M. *«recommenderlab: A Framework for Developing and Testing Recommendation Algorithms»* // The Intelligent Data Analysis Lab at SMU. – 2011. – URL: <http://lyle.smu.edu/IDA/recommenderlab/>.
12. HIKMAWATIE., MAULIDEVIN.U., SURENDRO K. *Minimum threshold determination method based on dataset characteristics in association rule mining* // J Big Data. – 2021. – Vol. 8. – P. 146. DOI: <https://doi.org/10.1186/s40537-021-00538-3>.
13. KVIYATKOVSKAYA I.Y., SHURSHEV V.F., FRENKEL M.B. *Methodology of a support of making management decisions for poorly structured problems* // Communications in

Computer and Information Science. – 2015. – Vol. 535. –  
P. 278–291.

14. <https://archive.ics.uci.edu/ml/datasets/lenses> – набор данных *Lenses (Lenses Data Set)*.

## **DEVELOPMENT OF A RECOMMENDATION MODEL TO SUPPORT DECISION-MAKING WHEN THE USER CHOOSES PRODUCTS**

**Irina Kvyatkovskaya**, Astrakhan State Technical University, Astrakhan, Doctor of technical sciences, professor (i.kvyatkovskaya@astu.org).

**Vo Thi Huyen Trang**, Astrakhan State Technical University, Astrakhan (vthtrang@mail.ru)

**Tran Quoc Toan**, Astrakhan State Technical University, Astrakhan (hoaiivan219@mail.ru)

*Abstract: Recommender systems are used to predict user preferences for a particular product or service, and to recommend suitable products or services to the user. Many of the methods used in data mining, related to classification or the construction of association rules, are used in recommender systems. This article proposes a new recommender model that combines association rules and statistical implication index measures. In the proposed model, support and confidence measures are used to create association rules, and the statistical implication index measure is used to filter the set of rules and rank recommendations. The proposed model and algorithms are used to build a recommendation result based on a known data set.*

**Keywords:** recommender system, recommender model, statistical implication analysis, statistical implication index, algorithm Apriori, association rules.

УДК 004.67

ББК 32.972.1

DOI: 10.25728/ubs.2023.105.6

*Статья представлена к публикации  
членом редакционной коллегии Р.М. Нижегородцевым.*

*Поступила в редакцию 29.03.2023.*

*Опубликована 30.09.2023.*