

Научная статья  
УДК 004.01:[651.4/.9]  
<https://doi.org/10.24143/2072-9502-2023-2-116-124>  
EDN UHNZRL

---

## **Модель и алгоритм поддержки принятия решения по выбору продуктов для рекомендации пользователю на основе метода анализа статистической импликации**

---

*Ирина Юрьевна Квятковская<sup>1✉</sup>, Чанг Во Тхи Хуен<sup>2</sup>, Тоан Чан Куок<sup>3</sup>*

<sup>1, 2</sup>*Астраханский государственный технический университет,  
Астрахань, Россия, [i.kvyatkovskaya@astu.org](mailto:i.kvyatkovskaya@astu.org)*

<sup>3</sup>*Каспийский институт морского и речного транспорта,  
филиал Волжского государственного университета водного транспорта,  
Астрахань, Россия*

---

**Аннотация.** Рассмотрены вопросы анализа информации, с которой сталкивается потребитель, осуществляя выбор продуктов и услуг. Проблемой является извлечение полезной информации, позволяющей предложить пользователю новые товары и услуги в зависимости от его предпочтений. Данную проблему локализуют рекомендательные системы, ориентированные на использование методов интеллектуального анализа данных, таких как классификация, кластеризация, анализ ассоциативных правил – метод машинного обучения, обнаруживающий связи между переменными в базах данных. По сравнению с другими методами преимущество рекомендательного метода на основе ассоциативных правил заключается в прозрачности – метод может продемонстрировать пользователю механизм вывода, использованный для принятия решений. В рекомендательных системах, основанных на ассоциативных правилах, используются две меры, широко применяемые для оценки наборов элементов и создания наборов правил ассоциации, – это мера поддержки (support) и мера достоверности (confidence). Однако чтобы получить более качественные рекомендации, качество ассоциативных правил и способ ранжирования предложений следует измерять с помощью некоторой объективной меры. Разработаны модель и алгоритм поддержки принятия решения по выбору продуктов для рекомендации пользователю на основе метода анализа статистической импликации. В предлагаемых решениях меры поддержки и достоверности используются для создания ассоциативных правил; мера интенсивности статистической импликации используется для фильтрации набора правил и ранжирования рекомендаций.

**Ключевые слова:** поиск, пользователь, продукт, рекомендательная система, рекомендательная модель, мера интенсивности статистической импликации, алгоритм Apriori, ассоциативное правило

**Для цитирования:** Квятковская И. Ю., Во Тхи Хуен Чанг, Чан Куок Тоан. Модель и алгоритм поддержки принятия решения по выбору продуктов для рекомендации пользователю на основе метода анализа статистической импликации // Вестник Астраханского государственного технического университета. Серия: Управление, вычислительная техника и информатика 2023. № 2. С. 116–124. <https://doi.org/10.24143/2072-9502-2023-2-116-124>. EDN UHNZRL.

Original article

---

## **Model and algorithm for supporting decision on selection of products for recommendation to user based on analysis of statistical implication**

---

*Irina Yu. Kvyatkovskaya<sup>1✉</sup>, Trang Vo Thi Huyen<sup>2</sup>, Toan Tran Quoc<sup>3</sup>*

<sup>1, 2</sup>*Astrakhan State Technical University,  
Astrakhan, Russia, [i.kvyatkovskaya@astu.org](mailto:i.kvyatkovskaya@astu.org)*

<sup>3</sup>*Caspian Institute of Sea and River Transport,  
branch of the Volga State University of Water Transport,  
Astrakhan, Russia*

---

**Abstract.** The article considers the issues of analyzing data that the consumer encounters when choosing products and services. The problem is extracting useful information that allows offering the user new products and services depending on his preferences. This problem is localized by recommender systems focused on using the data mining methods, such as classification, clustering, analysis of association rules - a machine learning method that detects relationships between variables in databases. Compared to other methods, the advantage of the association rule-based recommender method is its transparency: the method can show the user the inference mechanism used to make decisions. Association rule-based recommender systems use two measures that are widely popular and evaluate sets of elements and create sets of association rules, the support measure and the confidence measure. However, to get better recommendations, the quality of association rules and the way sentence ranking should be measured by some objective measure. There has been developed a model and a decision support algorithm for the choice of products for recommendations to the user based on the statistical implication analysis method. In the proposed solutions, support and credibility measures are used to create association rules; a measure of the intensity of the statistical implication is used to filter the set of rules and to rank the recommendations.

**Keywords:** search, user, product, recommender system, recommender model, statistical implication intensity, Apriori algorithm, association rule

**For citation:** Кыватковская И. Ю., Во Thi Huyen Trang, Tran Quoc Toan. Model and algorithm for supporting decision on selection of products for recommendation to user based on analysis of statistical implication. *Vestnik of Astrakhan State Technical University. Series: Management, computer science and informatics.* 2023;2:116-124. (In Russ.). <https://doi.org/10.24143/2073-5529-2023-2-116-124>. EDN UHNZRL.

## Введение

Современные поисковые системы сталкиваются с растущими вызовами: очень сложно найти полезную информацию для принятия решения при большом количестве вариантов за короткое время. Тенденция перехода от информационного поиска к информационно-рекомендательным советам осуществляется весьма быстро. Рекомендательная система стала крайне необходимым и широко используемым инструментом в сфере информационных технологий, коммерции, электронных услуг (AliExpress, Amazon, Pandora, Netflix, Walmart и т. д.). Цель рекомендательной системы состоит в том, чтобы отфильтровать полезную информацию из большого количества информации для возможности предсказания рейтинга, который пользователь присвоит элементу-продукту, и рекомендовать подходящие элементы [1–3].

Среди алгоритмов рекомендательных систем одним из самых успешных является совместная фильтрация. В то же время не существует внутренних проблем при выработке рекомендаций, если считать, что взаимное влияние пары пользователей симметрично [3]. Однако на практике роли и взаимодействия между двумя пользователями часто асимметричны. Это может создать определенную предвзятость в рекомендациях, т. к. более опытный пользователь, например эксперт, входящий в систему, имеет большее влияние на другого, а другой (менее опытный 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  $A$ , то  $B$ » между ними, опре-

деляя степень соответствия на основе статистических данных о взаимозависимостях между пользователями и выбранными продуктами [4]. В данном случае импликативное отношение рассматривается как отношение, основанное на знаниях о связях между пользователем и элементами данных, необходимых для выработки рекомендации. Таким образом, использование метода асимметричного подхода представляет большой интерес для минимизации систематической ошибки из-за указанного выше различия в результатах рекомендаций. Применение метода анализа статистической импликации в области рекомендательных систем необходимо для учета асимметричного влияния пользователей и оценки импликативной связи между пользователем и продуктами на практике.

## Определение меры интенсивности статистической импликации

Метод анализа статистической импликации [4, 5] – это метод анализа данных, который 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  $a$ ». Предположим, что набор  $E$  имеет  $n$  объектов или индивидов, описывается конечным набором бинарных переменных (атрибутов). 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  $n_{\bar{A}\bar{B}} = |A \cap \bar{B}|$  – это число объектов, которые удо-

влетворяют 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

$$\lambda = \frac{n_A \cdot n_{\bar{B}}}{n}.$$

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

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

где  $Pr$  – вероятность случайной величины.

Рассмотрим случай  $n_{\bar{B}} \neq 0$ . Тогда согласно расп-

$$\varphi(a, b) = 1 - Pr(Q(A, \bar{B}) \leq q(A, \bar{B})) = \begin{cases} \frac{1}{2\pi} \int_{q(A, \bar{B})}^{\infty} e^{-\frac{t^2}{2}} dt, & \text{если } n_B \neq n; \\ 0 & \text{другие;} \end{cases}$$

$$\varphi(a, b) = 1 - \sum_{k=\max(0, n_A - n_B)}^{n_{\bar{A}\bar{B}}} \frac{C_{n_B}^{n_A - k} C_{n - n_B}^k}{C_n^{n_A}},$$

где  $k$  – целое число.

Эта мера используется для определения контр-примеров  $n_{\bar{A}\bar{B}}$  в наборе  $E$ . Значение меры интенсивности статистической импликации  $\varphi(a, b)$  принимается с заданным порогом  $\alpha$ , если  $\varphi(a, b) \geq 1 - \alpha$ .

### Рекомендательная модель на основе меры интенсивности статистической импликации

Рекомендательная модель совместной фильтрации при поддержке принятия решений при выборе продуктов пользователем может быть представлена в виде следующей совокупности объектов:

$$RS = \{U, I, R, F\},$$

где  $U = \{u_1, u_2, \dots, u_n\}$  – набор из  $n$  пользователей;

$I = \{i_1, i_2, \dots, i_m\}$  – набор из  $m$  продуктов;

$R = \{r_{j,k}\}_{j=1..n}^{k=1..m}$  – матрица рейтингов пользователей,

где  $r_{j,k}$  – значение рейтинга пользователя  $u_j$  для продукта  $i_k$ ;  $F: U \times I \times R \rightarrow I_{u_a}$  – это вычислительная функция для поиска набора продуктов

$I_{u_a} = \{i_{u_a,1}, i_{u_a,2}, \dots, i_{u_a,N}\}$  для рекомендации пользователей  $u_a$ .

ределению Пуассона вероятность случайной величины  $|X \cap \bar{Y}|$  определяется следующим образом:

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

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

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

Это значение называется индексом статистической значимости.

Мера интенсивности статистической импликации  $\varphi(a, b)$  правила  $a \rightarrow b$  определяется формулой

Основными компонентами рекомендательных методов являются меры, используемые для поиска похожих пользователей, поиска похожих продуктов, выявления отношений с высоким уровнем вовлеченности или ранжирования рекомендаций и т. д. В предлагаемой модели показатели «поддержка» (англ. Support) и «достоверность» (англ. Confidence) используются для создания набора правил, а мера интенсивности статистической импликации используется для ранжирования набора правил. Основным входом предлагаемой модели является рейтинговая матрица, в которой хранится важная информация, генерируемая в процессе взаимодействия пользователя с системой, например рейтинг пользователя для просмотренных фильмов или прослушанных песен. Рейтинговая матрица может иметь пустые ячейки (отсутствуют рейтинги), если пользователь  $u_j$  не ставил оценку продукту  $i_k$  или не знает об этом.

Процесс работы предлагаемой модели показан на рис. 1.

На основе модели разработана методика поддержки принятия решения по выбору продуктов для рекомендации пользователю. В методике выделены 5 основных этапов.

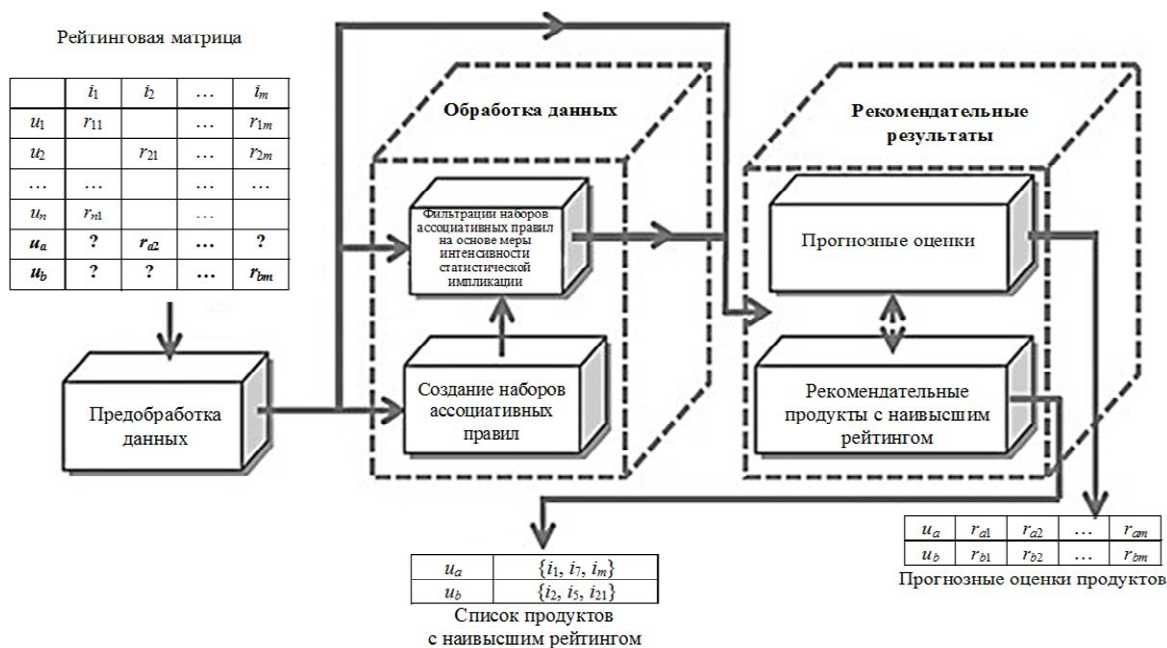


Рис. 1. Основные этапы обработки рекомендательной модели на основе меры интенсивности статистической импликации

Fig. 1. Main stages of processing the recommender model based on the measure of intensity of the statistical implication

Этап 1: предварительная обработка данных. Этот этап состоит из двух подзадач: 1) выбор соответствующих данных (в наборе данных) определением минимального количества элементов, оцененных каждым пользователем, и минимального количества пользователей, выполняющих обзор каждого элемента; 2) приведение к бинарному виду выбранного набора данных, если это не двоичные данные. Результатом этого этапа является набор данных (dataset).

Этап 2: создание набора ассоциативных правил *ruleset* из набора данных *dataset*. Этот этап состоит из двух подзадач: 1) выбор порога для 2-х показателей (поддержка и достоверность); 2) генерация набора ассоциативных правил на основе пороговых значений показателей (поддержка и достоверность) с использованием алгоритма Apriori [6]. Этот набор правил включает правила с максимальной длиной  $l$  и минимальной 2. Если правило имеет длину  $l$ , то оно будет иметь  $l-1$  элементов слева и  $l$  элемент справа.

Этап 3: создание набора правил фильтрацией набора правила *ruleset* на основе меры интенсивности статистической импликации. Этот этап состоит из трех подзадач: 1) представление каждого правила набором из 4-х значений  $\{n, n_a, n_b, n_{ab}\}$ ; 2) вычисление значения меры интенсивности статистической импликации каждого правила; 3) выбор соответствующего порога меры интенсивности статистической импликации для уменьшения набора правил.

Этап 4: поиск наиболее релевантного продукта для пользователей в наборе данных *reset* (это набор данных из  $p$  пользователей с отсутствием некоторых рейтингов). Этот этап состоит из двух подзадач: 1) для каждого пользователя найти правила, левая часть которых является подмножеством множества продуктов, которые оценены этим пользователем; 2) ранжировать (сортировать) отфильтрованные правила в соответствии со значением меры интенсивности статистической импликации и найти  $N$  предметов с наивысшим рейтингом.

Этап 5: прогноз рейтинга для отсутствующих оценок. Этот этап состоит из двух подзадач: 1) для каждой отсутствующей оценки в строке  $i$  и столбце  $j$  найти правила, у которых левая часть является подмножеством набора элементов, оцененных пользователем  $i$ , а правая часть является элементом  $j$ ; 2) ранжировать (сортировать) отфильтрованные правила в соответствии со значением меры интенсивности статистической импликации, затем присвоить отсутствующему рейтингу значение 0 или 1 на основе определенного порога.

В данной статье существенно раскрываются этапы 2, 5.

Детализируем некоторые сложносоставные этапы. Для реализации этапа 2 предложен алгоритм создания набора ассоциативных правил из набора обучающих данных. Вначале алгоритм генерирует наборы из 1 элемента, набора из 2 элементов, ...

Квятковская И. Ю., Во Тхи Хуен Чанг, Чан Куок Тоан. Модель и алгоритм поддержки принятия решения по выбору продуктов для рекомендации пользователю на основе метода анализа статистической импликации

набора из  $k$  элементов. Затем для каждого набора из  $k$  элементов генерируются непустые подмножества. Наконец, из этих подмножеств создаются ассоциативные правила на основе порогового значения меры интенсивности статистической импликации.

Этот порог определяется на основе экспериментальных результатов конкретных наборов данных. Алгоритм формирования ассоциативных правил, основанный на мере интенсивности статистической импликации, представлен на рис. 2.

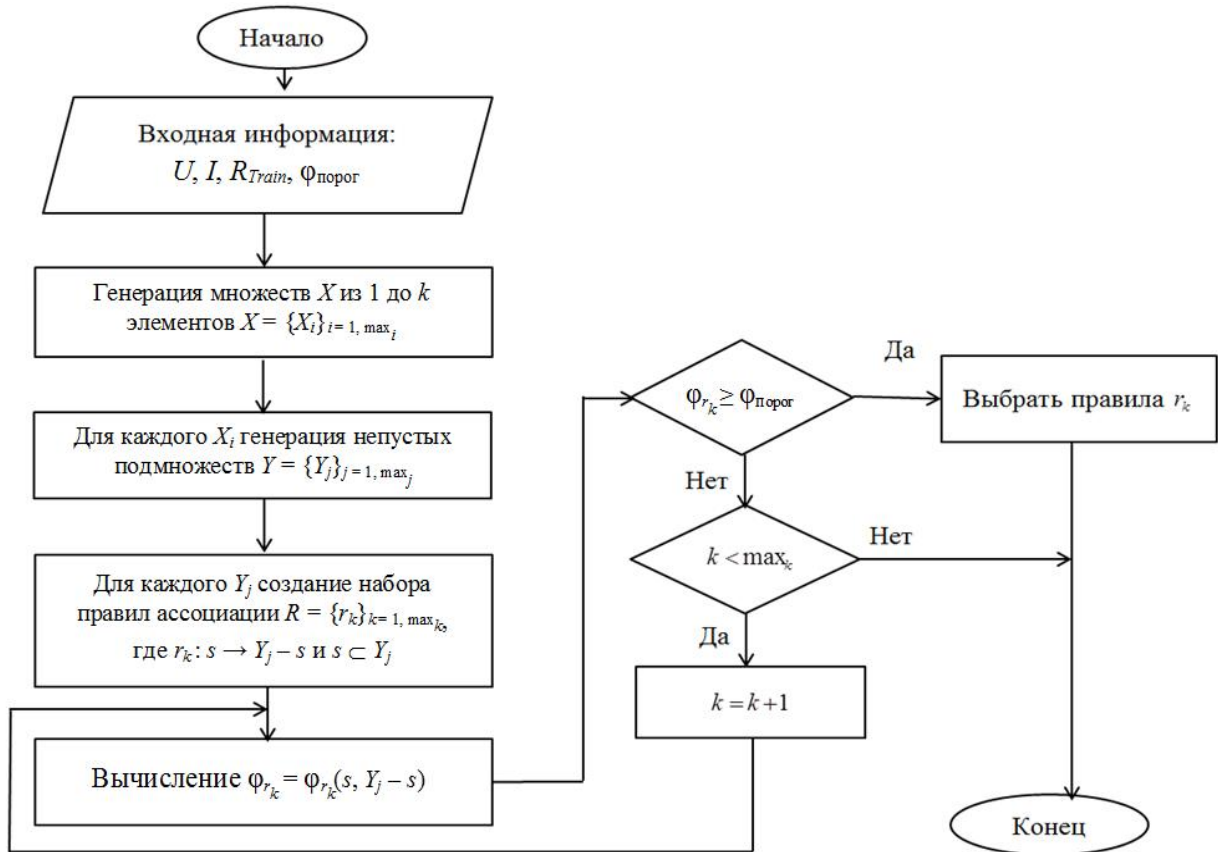


Рис. 2. Алгоритм формирования ассоциативных правил, основанный на мере интенсивности статистической импликации

Fig. 2. Algorithm for developing association rules based on the measure of intensity of the statistical implication

**Псевдокод алгоритма формирования правил ассоциации, основанный на мере интенсивности статистической импликации.**

*Входные данные:* набор обучающих данных и пороговое значение меры интенсивности статистической импликации.

*Выходные данные:* набор ассоциативных правил, основанный на мере интенсивности статистической импликации.

*Начало.*

*Шаг 1.* Сгенерировать наборы из 1 до  $k$  элементов:

<Создать набор из 1 элемента обучающих данных>;

<Создать набор из 2 элементов из набора из 1 элемента>;

...

<Создать набор из  $k$  элементов из набора из  $k - 1$  элементов>.

*Шаг 2.* Создать непустые подмножества для каждого набора кандидатов. Для каждого набора кандидатов  $I$  выполнить следующую операцию:

<Создать непустые подмножества  $s \subset I$ >.

*Шаг 3.* Сгенерировать ассоциативные правила с порогом меры интенсивности статистической импликации. Для каждого непустого подмножества  $s \subset I$  выполнить следующую операцию:

<Создать ассоциативное правило  $r: s \rightarrow (I - s)$ >.

Если значение меры интенсивности статистической импликации ассоциативного правила  $r$  больше, чем порог меры интенсивности статистической импликации, то:

<Выбрать ассоциативное правило  $r$  для рекомендательной модели>.

*Конец.*

**Пример 1.** Использовать рейтинговую матрицу для 2 пользователей и 4 продуктов следующим образом (табл. 1).

Таблица 1

Table 1

**Рейтинговая матрица  
для 2 пользователей и 4 продуктов  
Rating matrix for 2 users and 4 products**

Пользователь \ Продукт	$i_1$	$i_2$	$i_3$	$i_4$
$u_1$	0	4	4	1
$u_2$	0	0	4	1

Для создания ассоциативных правил матрица рейтинга пользователя преобразуется в разреженную бинарную матрицу со следующей структурой (табл. 2).

Таблица 2

Table 2

**Разреженная бинарная матрица рейтинга  
Sparse binary rating matrix**

$i_1 = 0$	$i_2 = 4$	$i_2 = 0$	$i_3 = 4$	$i_4 = 1$
1	1	0	1	1
1	0	1	1	1

Из разреженной бинарной матрицы определяются одноэлементные подмножества:

$$\{i_1 = 0\}, \{i_2 = 4\}, \{i_2 = 0\}, \{i_3 = 4\}, \{i_4 = 1\}.$$

Из 1-элементных подмножеств генерируются 2-элементные подмножества:

$$\{i_1 = 0, i_2 = 4\}, \{i_1 = 0, i_2 = 0\}, \{i_1 = 0, i_3 = 4\}, \\ \{i_1 = 0, i_4 = 1\}, \{i_2 = 4, i_2 = 0\}, \{i_2 = 4, i_3 = 4\}, \\ \{i_2 = 4, i_4 = 1\}, \{i_2 = 0, i_3 = 4\}, \{i_2 = 0, i_4 = 1\}, \\ \{i_3 = 4, i_4 = 1\}.$$

Из 2-элементных подмножеств генерируются трехэлементные подмножества:

$$\{i_1 = 0, i_2 = 4, i_2 = 0\}, \{i_1 = 0, i_2 = 4, i_3 = 4\}, \\ \{i_1 = 0, i_2 = 4, i_4 = 1\}, \{i_1 = 0, i_2 = 0, i_3 = 4\}, \\ \{i_1 = 0, i_2 = 0, i_4 = 1\}, \{i_1 = 0, i_3 = 4, i_4 = 1\}, \\ \{i_2 = 4, i_2 = 0, i_3 = 4\}, \{i_2 = 4, i_2 = 0, i_4 = 1\}, \\ \{i_2 = 4, i_3 = 4, i_4 = 1\}, \{i_2 = 0, i_3 = 4, i_4 = 1\}.$$

Из 3-элементных подмножеств генерируются 4-элементные подмножества:

$$\{i_1 = 0, i_2 = 4, i_2 = 0, i_3 = 4\}, \\ \{i_1 = 0, i_2 = 4, i_2 = 0, i_4 = 1\}, \\ \{i_1 = 0, i_2 = 4, i_3 = 4, i_4 = 1\}, \\ \{i_1 = 0, i_2 = 0, i_3 = 4, i_4 = 1\}, \\ \{i_2 = 4, i_2 = 0, i_3 = 4, i_4 = 1\}.$$

Из 4-элементных подмножеств генерируются 5-элементные подмножества:

$$\{i_1 = 0, i_2 = 4, i_2 = 0, i_3 = 4, i_4 = 1\}.$$

На основе полученных наборов приступаем к генерации ассоциативных правил с порогом ме-

ры интенсивности статистической импликации  $\phi(a, b) \geq 0,3$  и получаем 36 ассоциативных правил:

$$\{i_1 = 0\} \Rightarrow \{i_2 = 4\}, \{i_2 = 4\} \Rightarrow \{i_1 = 0\}, \{i_1 = 0\} \Rightarrow \{i_4 = 1\}, \\ \{i_4 = 1\} \Rightarrow \{i_1 = 0\}, \{i_2 = 4\} \Rightarrow \{i_4 = 1\}, \{i_4 = 1\} \Rightarrow \{i_2 = 4\}, \\ \{i_1 = 2\} \Rightarrow \{i_2 = 0\}, \{i_2 = 0\} \Rightarrow \{i_1 = 2\}, \{i_1 = 2\} \Rightarrow \{i_4 = 1\}, \\ \{i_4 = 1\} \Rightarrow \{i_1 = 2\}, \{i_2 = 0\} \Rightarrow \{i_4 = 1\}, \{i_4 = 1\} \Rightarrow \{i_2 = 0\}, \\ \{i_1 = 0, i_2 = 4\} \Rightarrow \{i_4 = 1\}, \{i_1 = 0, i_4 = 1\} \Rightarrow \{i_2 = 4\}, \\ \{i_2 = 4, i_4 = 1\} \Rightarrow \{i_1 = 0\}, \{i_1 = 0, i_3 = 4\} \Rightarrow \{i_2 = 4\}, \\ \{i_2 = 4, i_3 = 4\} \Rightarrow \{i_1 = 0\}, \{i_1 = 0, i_3 = 4\} \Rightarrow \{i_4 = 1\}, \\ \{i_3 = 4, i_4 = 1\} \Rightarrow \{i_1 = 0\}, \{i_2 = 4, i_3 = 4\} \Rightarrow \{i_4 = 1\}, \\ \{i_3 = 4, i_4 = 1\} \Rightarrow \{i_2 = 4\}, \{i_1 = 2, i_2 = 0\} \Rightarrow \{i_4 = 1\}, \\ \{i_1 = 2, i_4 = 1\} \Rightarrow \{i_2 = 0\}, \{i_2 = 0, i_4 = 1\} \Rightarrow \{i_1 = 2\}, \\ \{i_1 = 2, i_3 = 4\} \Rightarrow \{i_2 = 0\}, \{i_2 = 0, i_3 = 4\} \Rightarrow \{i_1 = 2\}, \\ \{i_1 = 2, i_3 = 4\} \Rightarrow \{i_4 = 1\}, \{i_3 = 4, i_4 = 1\} \Rightarrow \{i_1 = 2\}, \\ \{i_2 = 0, i_3 = 4\} \Rightarrow \{i_4 = 0\}, \{i_3 = 4, i_4 = 1\} \Rightarrow \{i_2 = 0\}, \\ \{i_1 = 0, i_2 = 4, i_3 = 4\} \Rightarrow \{i_4 = 1\}, \{i_1 = 0, i_3 = 4, i_4 = 1\} \Rightarrow \{i_2 = 4\}, \\ \{i_2 = 4, i_3 = 4, i_4 = 1\} \Rightarrow \{i_1 = 0\}, \{i_1 = 2, i_2 = 0, i_3 = 4\} \Rightarrow \{i_4 = 1\}, \\ \{i_1 = 2, i_3 = 4, i_4 = 1\} \Rightarrow \{i_2 = 0\}, \{i_2 = 0, i_3 = 4, i_4 = 1\} \Rightarrow \{i_1 = 2\}$$

Детализируем этап 5. Разработан алгоритм поддержки принятия решений по выбору продуктов для рекомендации пользователю на основе меры интенсивности статистической импликации.

Пусть  $U = \{u_1, u_2, \dots, u_n\}$  – набор из  $n$  пользователей,  $I = \{i_1, i_2, \dots, i_m\}$  – набор из  $m$  продуктов,  $R_{Train} = \{r_{j, k}\}$  – набор обучающих данных, где  $r_{j, k}$  – значение рейтинга пользователя  $u_j$  для продукта  $i_k$  в наборе обучающих данных;  $R_{Test} = \{r_{i, l}\}$  – набор тестовых данных, где  $r_{i, l}$  – значение рейтинга пользователя  $u_i$  для продукта  $i_l$  в наборе тестовых данных;  $RU = \{r_1, r_2, \dots, r_t\}$  – набор ассоциативных правил, сгенерированный из набора обучающих данных на основе меры интенсивности статистической импликации;  $MATRIX_{Logical} = \{b_{t, i}\}$  – логическая матрица, где  $b_{t, i}$  – логическое значение проверки соответствия между правилом  $r_t$  и пользователем  $u_i$ : если пользователь  $u_i$  имеет рейтинги продуктов в левой части правила  $r_t$ , то  $b_{t, i} = TRUE$ , иначе  $b_{t, i} = FALSE$ ,  $RU_{u_a} = \{r_{u_a, 1}, r_{u_a, 2}, \dots, r_{u_a, k}\}$  – набор ассоциативных правил, выбранных для пользователя  $u_a$ . Результатом рекомендации для пользователя  $u_a$  является набор продуктов, которые принадлежат правой части набора правил  $RU_{u_a}$ , и эти продукты не оценены пользователем  $u_a$ :  $I_{u_a} = \{i_{u_a, 1}, i_{u_a, 2}, \dots, i_{u_a, N}\}$ .

Чтобы определить набор ассоциативных правил для каждого пользователя в наборе тестовых данных, вначале алгоритм использует данные о количестве ассоциативных правил, сгенерированных из набора обучающих данных, и количестве пользователей в наборе тестовых данных для генерации логической матрицы, состоящей из  $t$  строк,  $l$  столбцов ( $t$  – количество правил,  $l$  – количество пользователей). Матрица создается следующим образом: если пользователь  $u_i$  имеет рейтинги товаров в левой части правила  $r_t$ , то ячейке  $(t, i)$  матрицы присваивается значение «TRUE» ( $b_{t, i} = TRUE$ ), в противном случае ей присваивается значение «FALSE» ( $b_{t, i} = FALSE$ ). Затем для каждого столбца  $i$ , если значение ячейки  $(t, i)$  равно



Квятковская И. Ю., Во Тхи Хуен Чанг, Чан Куок Тоан. Модель и алгоритм поддержки принятия решения по выбору продуктов для рекомендации пользователю на основе метода анализа статистической импликации

«TRUE», выбирается соответствующее ассоциативное правило в строке  $t$  для пользователя  $i$ . Наконец, выбранные правила сортируются для каждого пользователя, для которого вырабатываются рекомендации, в порядке убывания величины меры интенсивности статистической импликации. В итоге выбираются  $N$  продуктов из правой части правил с наибольшим значением, для которых пользователь не имеет рейтингов для предоставления рекомендаций.

**Псевдокод алгоритма поддержки принятия решений по выбору продуктов для рекомендации пользователю на основе меры интенсивности статистической импликации.**

**Входные данные:** набор ассоциативных правил, генерируемый из набора обучающих данных, набор тестовых данных.

**Выходные данные:** рекомендации для каждого пользователя в тестовом наборе данных.

**Начало.**

**Шаг 1.** Создание логической матрицы на основе набора ассоциативных правил и набора тестовых данных:

$t =$  <число ассоциативных правил>;  
 $l =$  <число пользователей набора тестовых данных>;  
 $MATRIX_{Logical} = \{b_{ij}\}$ , где  $i = \{1, 2, \dots, t\}$ ,  $j = \{1, 2, \dots, l\}$ .

Для каждого выбранного ассоциативного правила выполнить:

Если  $((\text{Набор продуктов в левой части правила в строке } i) \cap (\text{Набор продуктов, которые не оценены пользователем в столбце } j)) \neq \emptyset$ , то:

$MATRIX_{Logical}[i, j] = \text{"TRUE"}$ .

В противном случае:  $MATRIX_{Logical}[i, j] = \text{"FALSE"}$ .

**Шаг 2.** Выбор ассоциативных правил для каждого пользователя, с которым необходимо проконсультироваться.

Для каждого пользователя в тестовом наборе выполнить:

Если  $(MATRIX_{Logical}[j, i] = \text{"TRUE"})$ , то <Выбрать ассоциативное правило в строке  $i$  для пользователя в столбце  $j$ >.

**Шаг 3.** Выбрать продукты, по которым требуется проконсультироваться для каждого пользователя.

Для каждого пользователя в тестовом наборе выполнить:

<Сортировать правила в порядке убывания величины меры интенсивности статистической импликации>;

<Выбрать продукты справа от правила с наибольшим значением меры интенсивности статистической импликации, которые не оценены пользователем  $j$ , чтобы рекомендовать пользователю  $j$ >.

**Конец.**

**Пример 2.** Чтобы более четко увидеть этапы реализации модели, предположим, что система рекомендует пользователям выбрать 8 продуктов (от  $i_1$  до  $i_8$ ), система имеет 10 ранжированных пользователей (от  $u_1$  до  $u_{10}$ ). Рейтинги для продуктов, представленные бинарной матрицей, показаны в табл. 3.

Таблица 3

Table 3

Матрица рейтингов продуктов

Product rating matrix

Пользователь \ Продукт	$i_1$	$i_2$	$i_3$	$i_4$	$i_5$	$i_6$	$i_7$	$i_8$
$u_1$	0	1	1	1	1	1	0	0
$u_2$	1	0	0	0	1	1	0	0
$u_3$	1	0	0	1	1	1	0	1
$u_4$	1	0	1	1	1	0	1	1
$u_5$	1	1	0	0	0	0	0	1
$u_6$	0	1	0	0	0	1	0	1
$u_7$	1	1	0	1	0	1	1	0
$u_8$	1	0	0	1	0	1	0	0
$u_9$	0	1	1	0	0	0	1	0
$u_{10}$	1	0	1	0	0	0	0	1

Данные делятся на два набора: обучающих данных (от  $u_1$  до  $u_7$ ); тестовых данных (от  $u_8$  до  $u_{10}$ ), каждый пользователь выбирает два продукта для рекомендации. Первым шагом является создание набора ассоциативных правил на основе меры интенсивности статистической импликации на наборе обучающих данных с порогом меры интен-

сивности статистической импликации  $\phi(a, b) \geq 0,5$ , т. е. 49 правил. Следующий шаг выполняется на основе набора ассоциативных правил и набора тестов для построения логической матрицы. Последний шаг – от логической матрицы до генерации результатов рекомендаций – представлен в табл. 4.

Таблица 4  
 Table 4

Наборы правил для каждого пользователя и продукты для рекомендации пользователю  
 Sets of rules for each user and products to be recommended to the user

Квыатковская И. Ю., Во Thi Нгуен Тамг, Триан Куок Тоан. Model and algorithm for supporting decision on selection of products for recommendation to user based on analysis of statistical implication

Пользователь	Ассоциативные правила	Значение меры интенсивности статистической импликации	Продукты для рекомендации
$u_8$	$\{i_4, i_5\} \Rightarrow \{i_3\}; \{i_1, i_4\} \Rightarrow \{i_7\}; \{i_3, i_4\} \Rightarrow \{i_5\}; \{i_4, i_8\} \Rightarrow \{i_5\};$ $\{i_1, i_5, i_8\} \Rightarrow \{i_4\}; \{i_1, i_4, i_8\} \Rightarrow \{i_5\}; \{i_1, i_4, i_5\} \Rightarrow \{i_8\}; \{i_4\} \Rightarrow \{i_3\};$ $\{i_4\} \Rightarrow \{i_7\}; \{i_4\} \Rightarrow \{i_5\}; \{i_1, i_3\} \Rightarrow \{i_7\}; \{i_3, i_4, i_8\} \Rightarrow \{i_7\};$ $\{i_4, i_7, i_8\} \Rightarrow \{i_3\}; \{i_1, i_3, i_8\} \Rightarrow \{i_7\}; \{i_1, i_7, i_8\} \Rightarrow \{i_3\}; \{i_4, i_5, i_7\} \Rightarrow \{i_3\};$ $\{i_1, i_3, i_5\} \Rightarrow \{i_7\}; \{i_1, i_5, i_7\} \Rightarrow \{i_3\}; \{i_1, i_3, i_4\} \Rightarrow \{i_7\}; \{i_2, i_4, i_5\} \Rightarrow \{i_3\};$ $\{i_2, i_5, i_6\} \Rightarrow \{i_3\}; \{i_1, i_2, i_4\} \Rightarrow \{i_7\}; \{i_1, i_2, i_6\} \Rightarrow \{i_7\};$ $\{i_3, i_4, i_5, i_8\} \Rightarrow \{i_7\}; \{i_4, i_5, i_7, i_8\} \Rightarrow \{i_3\}; \{i_1, i_3, i_5, i_8\} \Rightarrow \{i_7\};$ $\{i_1, i_5, i_7, i_8\} \Rightarrow \{i_3\}; \{i_1, i_3, i_4, i_8\} \Rightarrow \{i_7\}; \{i_1, i_4, i_7, i_8\} \Rightarrow \{i_3\};$ $\{i_1, i_3, i_4, i_5\} \Rightarrow \{i_7\}; \{i_1, i_4, i_5, i_7\} \Rightarrow \{i_3\}; \{i_2, i_4, i_5, i_6\} \Rightarrow \{i_3\};$ $\{i_1, i_2, i_4, i_6\} \Rightarrow \{i_7\}; \{i_1, i_3, i_4, i_5, i_8\} \Rightarrow \{i_7\}; \{i_1, i_4, i_5, i_7, i_8\} \Rightarrow \{i_3\}$	[0,63;0,51]	$i_3, i_7$
$u_9$	$\{i_3, i_4\} \Rightarrow \{i_5\}; \{i_3\} \Rightarrow \{i_5\}; \{i_3\} \Rightarrow \{i_4\}; \{i_7\} \Rightarrow \{i_4\};$ $\{i_3, i_5\} \Rightarrow \{i_4\}; \{i_1, i_7\} \Rightarrow \{i_4\}; \{i_1, i_3\} \Rightarrow \{i_7\}; \{i_3, i_4, i_8\} \Rightarrow \{i_7\};$ $\{i_4, i_7, i_8\} \Rightarrow \{i_3\}; \{i_1, i_3, i_8\} \Rightarrow \{i_7\}; \{i_1, i_7, i_8\} \Rightarrow \{i_3\};$ $\{i_4, i_5, i_7\} \Rightarrow \{i_3\}; \{i_1, i_5, i_7\} \Rightarrow \{i_3\}; \{i_1, i_3, i_4\} \Rightarrow \{i_7\}; \{i_2, i_4, i_5\} \Rightarrow \{i_3\};$ $\{i_2, i_5, i_6\} \Rightarrow \{i_3\}; \{i_1, i_2, i_4\} \Rightarrow \{i_7\}; \{i_1, i_2, i_6\} \Rightarrow \{i_7\};$ $\{i_3, i_4, i_5, i_8\} \Rightarrow \{i_7\}; \{i_4, i_5, i_7, i_8\} \Rightarrow \{i_3\}; \{i_1, i_3, i_5, i_8\} \Rightarrow \{i_7\};$ $\{i_1, i_5, i_7, i_8\} \Rightarrow \{i_3\}; \{i_1, i_3, i_4, i_8\} \Rightarrow \{i_7\}; \{i_1, i_4, i_7, i_8\} \Rightarrow \{i_3\};$ $\{i_1, i_3, i_4, i_5\} \Rightarrow \{i_7\}; \{i_1, i_4, i_5, i_7\} \Rightarrow \{i_3\}; \{i_2, i_4, i_5, i_6\} \Rightarrow \{i_3\};$ $\{i_1, i_2, i_4, i_6\} \Rightarrow \{i_7\}; \{i_1, i_3, i_4, i_5, i_8\} \Rightarrow \{i_7\}; \{i_1, i_4, i_5, i_7, i_8\} \Rightarrow \{i_3\};$ $\{i_3, i_8\} \Rightarrow \{i_7\}; \{i_7, i_8\} \Rightarrow \{i_3\}; \{i_5, i_7\} \Rightarrow \{i_3\}; \{i_3, i_5, i_8\} \Rightarrow \{i_7\};$ $\{i_5, i_7, i_8\} \Rightarrow \{i_3\}; \{i_2, i_5\} \Rightarrow \{i_3\}$	[0,58; 0,51]	$i_5, i_4$
$u_{10}$	$\{i_1, i_4\} \Rightarrow \{i_7\}; \{i_3, i_4\} \Rightarrow \{i_5\}; \{i_3\} \Rightarrow \{i_5\}; \{i_3\} \Rightarrow \{i_4\}; \{i_3, i_5\} \Rightarrow \{i_4\};$ $\{i_4, i_8\} \Rightarrow \{i_5\}; \{i_1, i_5, i_8\} \Rightarrow \{i_4\}; \{i_1, i_4, i_8\} \Rightarrow \{i_5\}; \{i_1, i_4, i_5\} \Rightarrow \{i_8\};$ $\{i_5, i_8\} \Rightarrow \{i_4\}; \{i_1, i_3\} \Rightarrow \{i_7\}; \{i_3, i_4, i_8\} \Rightarrow \{i_7\}; \{i_4, i_7, i_8\} \Rightarrow \{i_3\};$ $\{i_1, i_3, i_8\} \Rightarrow \{i_7\}; \{i_1, i_7, i_8\} \Rightarrow \{i_3\}; \{i_1, i_5, i_7\} \Rightarrow \{i_3\}; \{i_1, i_3, i_4\} \Rightarrow \{i_7\};$ $\{i_1, i_2, i_4\} \Rightarrow \{i_7\}; \{i_1, i_2, i_6\} \Rightarrow \{i_7\}; \{i_3, i_4, i_5, i_8\} \Rightarrow \{i_7\};$ $\{i_4, i_5, i_7, i_8\} \Rightarrow \{i_3\}; \{i_1, i_3, i_5, i_8\} \Rightarrow \{i_7\}; \{i_1, i_5, i_7, i_8\} \Rightarrow \{i_3\};$ $\{i_1, i_3, i_4, i_8\} \Rightarrow \{i_7\}; \{i_1, i_4, i_7, i_8\} \Rightarrow \{i_3\}; \{i_1, i_3, i_4, i_5\} \Rightarrow \{i_7\};$ $i_1, i_4, i_5, i_7 \Rightarrow \{i_3\}; \{i_1, i_2, i_4, i_6\} \Rightarrow \{i_7\}; \{i_1, i_3, i_4, i_5, i_8\} \Rightarrow \{i_7\};$ $\{i_1, i_4, i_5, i_7, i_8\} \Rightarrow \{i_3\}; \{i_3, i_8\} \Rightarrow \{i_7\}; \{i_7, i_8\} \Rightarrow \{i_3\}; \{i_3, i_8\} \Rightarrow \{i_7\};$ $\{i_5, i_7, i_8\} \Rightarrow \{i_3\}; \{i_1, i_3, i_5\} \Rightarrow \{i_7\}$	[0,63; 0,51]	$i_7, i_5$

Исходя из результатов рекомендации в табл. 4, обнаружены продукты для рекомендации пользователю  $u_8 - \{i_3, i_7\}$ , для пользователя  $u_9 - \{i_5, i_4\}$ , для пользователя  $u_{10} - \{i_7, i_5\}$ .

**Заключение**

Разработаны модель и алгоритм поддержки принятия решения по выбору продуктов для рекоменда-

ции пользователю на основе метода анализа статистической импликации. В модели и алгоритме показано, как определить набор ассоциативных правил на основе меры интенсивности статистической импликации, описаны шаги для построения рекомендательной модели на основе этого набора правил.

**Список источников**

1. Шуршев В. Ф., Кочкин Г. А., Кочкина В. Р. Модель системы поддержки принятия решений на основе рассуждений по прецедентам техника и информатика. 2013. № 2. С. 175–183.  
 2. Кыватковская И. Ю., Шуршев В. Ф., Френкел М. В. Methodology of a support of making management decisions for poorly structured problems // Communications in Computer and Information Science. 2015. V. 535. P. 278–291.  
 3. Нефедова Ю. С. Архитектура гибридной рекомендательной системы GEFEST (Generation–Expansion–

Filtering–Sorting–Truncation) // Системы и средства информатики. 2012. Т. 22, вып. 2. С. 176–196.  
 4. Gras R., Kuntz P. An overview of the Statistical Implicative Analysis (SIA) development // Statistical Implicative Analysis - Studies in Computational Intelligence. 2008. V. 127. P. 11–40.  
 5. Gras R., Kuntz P., Greffard N. Notion of Implicative Fields in Statistical Implicative Analysis // VIII Colloque International - VIII International Conference A.S.I. Analyse



Statistique Implicative - Statistical Implicative Analysis Radès (Tunisie) - Novembre 2015. P. 29–46.

6. Шитиков В. К., Мастицкий С. Э. Классификация, регрессия, алгоритмы Data Mining с использованием

R, 2017. URL: <https://github.com/ranalytics/data-mining> (дата обращения: 07.01.2019).

### References

1. Shurshev V. F., Kochkin G. A., Kochkina V. R. Model' sistemy podderzhki priniatiia reshenii na osnove rassuzhdenii po pretsedentam [Decision support system model based on case-based reasoning]. *Vestnik Astrakhanskogo gosudarstvennogo tekhnicheskogo universiteta. Seriya: Upravlenie, vychislitel'naiia tekhnika i informatika*, 2013, no. 2, pp. 175-183.

2. Kvyatkovskaya 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, pp. 278-291.

3. Nefedova Iu. S. Arkhitektura gibridnoi rekomendat'noi sistemy GEFEST (Generation–Expansion–Filtering–Sorting–Truncation) Architecture of the hybrid recommender system GEFEST (Generation–Expansion–

Filtering–Sorting–Truncation)]. *Sistemy i sredstva informatiki*, 2012, vol. 22, iss. 2, pp. 176-196.

4. Gras R., Kuntz P. An overview of the Statistical Implicative Analysis (SIA) development. *Statistical Implicative Analysis - Studies in Computational Intelligence*, 2008, vol. 127, pp. 11-40.

5. Gras R., Kuntz P., Greffard N. Notion of Implicative Fields in Statistical Implicative Analysis. *VIII Colloque International - VIII International Conference A.S.I. Analyse Statistique Implicative - Statistical Implicative Analysis Radès (Tunisie) - Novembre 2015*. Pp. 29-46.

6. Shitikov V. K., Mastitskii S. E. *Klassifikatsiia, regressiia, algoritmy Data Mining s ispol'zovaniem R, 2017* [Classification, regression, Data Mining algorithms using R, 2017]. Available at: <https://github.com/ranalytics/data-mining> (accessed: 07.01.2019).

Статья поступила в редакцию 01.02.2023; одобрена после рецензирования 30.03.2023; принята к публикации 24.04.2023  
The article is submitted 01.02.2023; approved after reviewing 30.03.2023; accepted for publication 24.04.2023

### Информация об авторах / Information about the authors

**Ирина Юрьевна Квятковская** – доктор технических наук, профессор; заведующий кафедрой высшей и прикладной математики; Астраханский государственный технический университет; [i.kvyatkovskaya@astu.org](mailto:i.kvyatkovskaya@astu.org)

**Чанг Во Тхи Хуен** – аспирант кафедры автоматизированных систем обработки информации и управления; Астраханский государственный технический университет; [vthtrang@mail.ru](mailto:vthtrang@mail.ru)

**Тоан Чан Куок** – кандидат технических наук; преподаватель кафедры судомеханических дисциплин; Каспийский институт морского и речного транспорта, филиал Волжского государственного университета водного транспорта; [hoaiivan219@mail.ru](mailto:hoaiivan219@mail.ru)

**Irina Yu. Kvyatkovskaya** – Doctor of Technical Sciences, Professor; Head of the Department of Higher and Applied Mathematics; Astrakhan State Technical University; [i.kvyatkovskaya@astu.org](mailto:i.kvyatkovskaya@astu.org)

**Trang Vo Thi Huyen** – Postgraduate Student of the Department of Automated Systems of Information Processing and Control; Astrakhan State Technical University; [vthtrang@mail.ru](mailto:vthtrang@mail.ru)

**Toan Tran Quoc** – Candidate of Technical Sciences; Lecturer of the Department of Ship-mechanical Disciplines; Caspian Institute of Sea and River Transport, branch of the Volga State University of Water Transport; [hoaiivan219@mail.ru](mailto:hoaiivan219@mail.ru)

