

ISSN 2708-2032  
e-ISSN 2708-2040



**INTERNATIONAL  
UNIVERSITY**

**INTERNATIONAL  
JOURNAL OF INFORMATION  
& COMMUNICATION TECHNOLOGIES**

---

**Volume 2, Issue 2  
June, 2021**

ҚАЗАҚСТАН РЕСПУБЛИКАСЫНЫҢ БІЛІМ ЖӘНЕ ҒЫЛЫМ МИНИСТРЛІГІ  
МИНИСТЕРСТВО ОБРАЗОВАНИЯ И НАУКИ РЕСПУБЛИКИ КАЗАХСТАН  
MINISTRY OF EDUCATION AND SCIENCE OF THE REPUBLIC OF KAZAKHSTAN



**INTERNATIONAL JOURNAL OF  
INFORMATION AND COMMUNICATION  
TECHNOLOGIES**

**МЕЖДУНАРОДНЫЙ ЖУРНАЛ  
ИНФОРМАЦИОННЫХ И  
КОММУНИКАЦИОННЫХ ТЕХНОЛОГИЙ**

**ХАЛЫҚАРАЛЫҚ АҚПАРАТТЫҚ ЖӘНЕ  
КОММУНИКАЦИЯЛЫҚ  
ТЕХНОЛОГИЯЛАР ЖУРНАЛЫ**

Том 2, Выпуск 2  
Июнь, 2021

Главный редактор – Ректор АО МУИТ, профессор, д.т.н.  
**Ускенбаева Р.К.**

Заместитель главного редактора – Проректор по НиМД, PhD, ассоц.профессор  
**Дайнеко Е.А.**

Отв. секретарь – PhD, ассоц.профессор, директор департамента по науке  
**Кальпеева Ж.Б.**

#### **ЧЛЕНЫ РЕДКОЛЛЕГИИ:**

**Отельбаев М. д.т.н., профессор, АО «МУИТ», Рысбайулы Б., д.т.н., профессор, АО «МУИТ», Куандыков А.А., д.т.н., профессор, АО «МУИТ», Синчев Б.К., д.т.н., профессор, АО «МУИТ», Дузбаев Н.Т., PhD, проректор по ЦИИ, АО «МУИТ», Ыдырыс А., PhD, заведующая кафедрой «МКМ», АО «МУИТ», Касымова А.Б., PhD, заведующая кафедрой «ИС», АО «МУИТ», Шильдибеков Е.Ж., PhD, заведующий кафедрой «ЭиБ», АО «МУИТ», Ипалакова М.Т., к.т.н., ассоц. профессор, заведующая кафедрой «КИИБ», АО «МУИТ», Айтмагамбетов А.З., к.т.н., профессор, АО «МУИТ», Амиргалиева С.Н., д.т.н., профессор, АО «МУИТ», Ниязгулова А.А., к.ф.н., заведующая кафедрой «МииК», АО «МУИТ», Молдагулова А.Н., к.т.н., ассоциированный профессор, АО «МУИТ», Джоламанова Б.Д., ассоциированный профессор, АО «МУИТ», Prof. Young Im Cho, PhD, Gachon University, South Korea, Prof. Michele Pagano, PhD, University of Pisa, Italy, Tadeusz Wallas, Ph.D., D.Litt., Adam Mickiewicz University in Poznań, Тихвинский В.О., д.э.н., профессор, МТУСИ, Россия, Масалович А., к.ф.-м.н., Президент Консорциума Инфорус, Россия, Lucio Tommaso De Paolis is the Research Director of the Augmented and Virtual Laboratory (AVR Lab) of the Department of Engineering for Innovation, University of Salento and the Responsible of the research group on “Advanced Virtual Reality Application in Medicine” of the DREAM, a multidisciplinary research laboratory of the Hospital of Lecce (Italy), Liz Bacon, Professor, Deputy Principal and Deputy Vice-Chancellor, Abertay University (Great Britain).**

Издание зарегистрировано Министерством информации и общественного развития Республики Казахстан. Свидетельство о постановке на учет № KZ82VPY00020475 от 20.02.2020 г.

Журнал зарегистрирован в Международном центре по регистрации сериальных изданий ISSN (ЮНЕСКО, г. Париж, Франция)

Выходит 4 раза в год.

#### **УЧРЕДИТЕЛЬ:**

**АО «Международный университет информационных технологий»**

ISSN 2708-2032 (print)  
ISSN 2708-2040 (online)

## СОДЕРЖАНИЕ

### РАЗРАБОТКА ПРОГРАММНОГО ОБЕСПЕЧЕНИЯ И ИНЖЕНЕРИЯ ЗНАНИЙ

**Бактаев А.Б., Мукажанов Н.К.**

Алгоритм решения задачи по исправлению опечаток в тексте, применяемый в поисковых системах с поддержкой казахского языка ..... 9

**Еркетаев Н.М., Мукажанов Н.К.**

Эффективное хранение неструктурированных данных ..... 19

**Сагадиев Р.Т., Шайкемелев Г.Т.**

Представление логической витрины данных в экосистеме Hadoop ..... 28

**Бейсенбек Е.Б., Дузбаев Н.Т.**

Современные способы взлома и защиты ПО ..... 33

**Найзабаева Л.К., Алашымбаев Б.А.**

Рекомендательная система для онлайн-магазинов с использованием машинного обучения ..... 38

**Мейрамбайулы Н., Дузбаев Н.Т.**

Мониторинг стационарных источников выбросов загрязняющих веществ г. Алматы ..... 47

### ИНФОКОММУНИКАЦИОННЫЕ СЕТИ И КИБЕРБЕЗОПАСНОСТЬ

**Айтмагамбетов А.З., Кулакаева А.Е., Койшыбай С.С., Жолшибек И.Ж.**

Исследование возможностей применения низкоорбитальных спутников для радиомониторинга в республике Казахстан ..... 54

**Кемельбеков Б.Ж., Полуанов М.**

Анализ метода бриллюэновской рефлектометрии в волоконно-оптических линиях связи ... 62

**Турбекова К.Ж.**

Анализ применения БПЛА в сетях связи при чрезвычайных ситуациях ..... 68

### ИНТЕЛЛЕКТУАЛЬНЫЕ СИСТЕМЫ

**Азанов Н.П., Хабиров Р.Р., Әміров У.Е.**

Конкурентная разведка и принятие решений с помощью машинного обучения для обеспечения промышленной безопасности ..... 75

**Джаныбекова С.Т., Толганбаева Г.А., Сарсембаев А.А.**

Распознавание говорящего с помощью глубокого обучения ..... 85

**Салерова Д.К., Сарсембаев А.А.**

Обзорная статья распознавания номерных знаков с использованием оптического распознавания символов ..... 93

**Салерова Д.К., Сарсембаев А.А.**

Исследование существующих методов классификации изображений ..... 100

**Оразалин А., Мурсалиев Д.Е., Сергазина А.С.**

Актуальные сверточные архитектуры нейронной сети для диагностики медицинских изображений ..... 115

**Әлімхан А.М.**

Прогнозирование результатов игры в баскетбол с использованием алгоритмов глубокого обучения ..... 112

<b><i>Адырбек Ж.А., Сатыбалдиева Р.Ж.</i></b> Анализ процессов планирования и решения проблем в логистике с помощью интеллектуальной системы .....	120
<b><i>Нурғалиев М.К., Алимжанова Л.М.</i></b> Геймификация в образовании .....	128

## **ЦИФРОВЫЕ ТЕХНОЛОГИИ В ЭКОНОМИКЕ И МЕНЕДЖМЕНТЕ**

<b><i>Алимжанова Л.М., Панарина А.В.</i></b> Внедрение сервисной системы IT-аутсорсинга .....	133
<b><i>Жұмабай Р.Ж., Алимжанова Л.М.</i></b> Управление процессами работы с поставщиками на основе ERP-стандартов — подход BPM .....	140
<b><i>Бердыкулова Г.М., Төлепбергенова Д.А.</i></b> Менеджмент университета: практика МУИТ .....	146
<b><i>Омарова А.Ш., Алимжанова Л.М., Таштамышева А.Э.</i></b> Исследование и разработка методов перехода традиционного маркетинга в цифровой формат .....	153

## CONTENTS

### SOFTWARE DEVELOPMENT AND KNOWLEDGE ENGINEERING

<b><i>Baktayev A.B., Mukazhanov N.K.</i></b> Algorithm for solving the problem of correcting typos with search engines supporting the Kazakh language .....	9
<b><i>Yerketayev N.M., Mukazhanov N.K.</i></b> Efficient storage of unstructured data .....	19
<b><i>Sagadiyev R.T., Shaikemelev G.T.</i></b> Representing a logical data mart in the Hadoop ecosystem .....	28
<b><i>Beisenbek Y.B., Duzbaev N.T.</i></b> Modern methods of hacking and protection software .....	33
<b><i>Naizabayeva L., Alashybayev B.A.</i></b> A recommendation system for online stores using machine learning .....	38
<b><i>Meirambaiuly N., Duzbaev N.T.</i></b> Monitoring of stationary sources of pollutant emissions in Almaty .....	47

### INFORMATION AND COMMUNICATION NETWORKS AND CYBERSECURITY

<b><i>Aitmagambetov A.Z., Kulakayeva A.E., Koishybai S.S., Zholshibek I.Z.</i></b> Study of the possibilities of using low-orbit satellites for radio monitoring in the Republic of Kazakhstan .....	54
<b><i>Kemelbekov B.J., Poluanov M.</i></b> Analysis of the brillouin reflectometry method in fiber-optic communication lines .....	62
<b><i>Turbekova K.Zh.</i></b> Analysis of the use of UAVs in emergency communication networks .....	68

### SMART SYSTEMS

<b><i>Azanov N.P., Khabirov R.R., Amirov U.E.</i></b> Competitive intelligence and decision-making algorithm using machine learning for industrial security .....	75
<b><i>Janybekova S.T., Tolganbayeva G.A., Sarsembayev A.A.</i></b> Speaker recognition using deep learning .....	85
<b><i>Salerova D.K., Sarsembayev A.A.</i></b> Review of license plate recognition using optical character recognition .....	93
<b><i>Salerova D.K., Sarsembayev A.A.</i></b> Research on the existing image classification methods .....	100
<b><i>Orazalin A., Mursaliyev D.E., Sergazina A.S.</i></b> Current convolutional neural network architectures for diagnosing medical images.....	105
<b><i>Alimkhan A.M.</i></b> Predicting basketball results using deep learning algorithms .....	112
<b><i>Adyrbek Zh.A., Satybaldiyeva R.Zh.</i></b> Analysis of the planning and problem-solving processes in logistics using an intelligent system .....	120
<b><i>Nurgaliyev M.K., Alimzhanova L.M.</i></b> Gamification in education .....	128

## DIGITAL TECHNOLOGIES IN ECONOMICS AND MANAGEMENT

*Alimzhanova L.M., Panarina A.V.*

Implementation of an IT outsourcing service system ..... 133

*Zhumabay R.Zh., Alimzhanova L.M.*

Supplier process management based on ERP standards: the BPM approach ..... 140

*Berdykulova G.M., Tolepbergenova D.A.*

University management: case study of IITU ..... 146

*Omarova A.Sh., Alimzhanova L.M., Tashtamysheva A.E.*

Research and development of methods for the transition of traditional marketing to digital  
format ..... 153

## МАЗМҰНЫ

### БАҒДАРЛАМАЛЫҚ ҚАМТАМАНЫ ӨЗІРЛЕУ ЖӘНЕ БІЛІМ ИНЖЕНЕРИЯСЫ

**Бактаев А.Б., Мукажанов Н.К.**

Қазақ тілін қолдайтын іздеу жүйелерінде қолданылатын мәтіндегі жаңылыстарды түзету бойынша есептерді шешу алгоритмі..... 9

**Еркетаев Н.М., Мукажанов Н.К.**

Құрылымсыз деректерді тиімді сақтау ..... 19

**Сагадиев Р.Т., Шайкемелев Г.Т.**

Надоор экожүйесінде логикалық деректер кесіндісін ұсыну ..... 28

**Бейсенбек Е.Б., Дузбаев Н.Т.**

Бағдарламалық жасақтаманы бұзудың және қорғаудың заманауи әдістері ..... 33

**Найзабаева Л., Алашыбаев Б.А.**

Машиналық оқытуды қолдану арқылы интернет-дүкендерге арналған ұсыныс жүйесі ..... 38

**Мейрамбайұлы Н., Дузбаев Н.Т.**

Алматы қаласы бойынша ластаушы заттар шығарындыларының стационарлық дереккөздеріне мониторинг жүргізу ..... 47

### АҚПАРАТТЫҚ ЖӘНЕ КОММУНИКАЦИЯЛЫҚ ЖЕЛІЛЕР ЖӘНЕ КИБЕРҚАУПСІЗДІК

**Айтмагамбетов А.З., Қулакаева А.Е., Койшыбай С.С., Жолшибек И.Ж.**

Қазақстан Республикасында радиомониторинг үшін төмен орбиталық спутниктерді қолдану мүмкіндіктерін зерттеу ..... 54

**Кемельбеков Б.Ж., Полуанов М.**

Талшықты-оптикалық байланыс желілеріндегі бриллюэн рефлектометрия әдісін талдау ... 62

**Турбекова К.Ж.**

Төтенше жағдайлар кезінде байланыс желілерінде ПҰА-ның қолданылуын талдау ..... 68

### ИНТЕЛЛЕКТУАЛДЫ ЖҮЙЕЛЕР

**Азанов Н.П., Хабиров Р.Р., Әміров У.Е.**

Өнеркәсіптік қауіпсіздікті қамтамасыз ету үшін машиналық оқытуды қолдана отырып, бәсекеге қабілеттілікті барлау және шешім қабылдау ..... 75

**Джаныбекова С.Т., Толғанбаева Г.А., Сарсембаев А.А.**

Терең оқыту арқылы сөйлеушіні тану ..... 85

**Салерова Д.К., Сарсембаев А.А.**

Таңбаларды оптикалық тануды пайдалану арқылы нөмірлер белгілерін тануға шолу мақаласы ..... 93

**Салерова Д.К., Сарсембаев А.А.**

Қолданыстағы бейнелерді жіктеу әдістерін зерттеу ..... 100

**Оразалин А., Мурсалиев Д.Е., Сергазина А.С.**

Медициналық кейіндік диагностикаға арналған конволюциялық жүйкелік желі архитектурасы ..... 105

**Әлімхан А.М.**

Терең оқыту алгоритмдерін қолдана отырып, баскетбол нәтижелерін болжау ..... 112

<b><i>Адырбек Ж.А., Сатыбалдиева Р.Ж.</i></b> Логистикадағы жоспарлау процестерін талдау және логистикадағы интеллектуалды жүйені қолдану арқылы мәселелерді шешу .....	120
<b><i>Нұрғалиев М.Қ., Алимжанова Л.М.</i></b> Білім беру саласындағы геймификация .....	128

## ЭКОНОМИКА ЖӘНЕ БАСҚАРУДАҒЫ САНДЫҚ ТЕХНОЛОГИЯЛАР

<b><i>Алимжанова Л.М., Панарина А.В.</i></b> IT-аутсорсингтің сервистік жүйесін енгізу .....	133
<b><i>Жұмабай Р.Ж., Алимжанова Л.М.</i></b> ERP стандарттарына негізделген жеткізушілермен жұмыс процесін басқару - BPM тәсілі .....	140
<b><i>Бердыкулова Г.М., Төлепбергенова Д.А.</i></b> Университетті басқару: ХАТУ практикасы .....	146
<b><i>Омарова А.Ш., Алимжанова Л.М., Таштамышева А.Э.</i></b> Дәстүрлі маркетингті цифрлық форматқа ауыстыру әдістерін зерттеу және әзірлеу .....	153

Найзабаева Л.К\*, Алашыбаев Б.А

Международный университет информационных технологий, Алматы, Казахстан

## РЕКОМЕНДАТЕЛЬНАЯ СИСТЕМА ДЛЯ ОНЛАЙН-МАГАЗИНОВ С ИСПОЛЬЗОВАНИЕМ МАШИННОГО ОБУЧЕНИЯ

**Аннотация.** Среди множества последних трендов интернет-маркетинга можно отметить рекомендательные системы. Рекомендательные системы — особые приложения, направленные на прогнозирование интересов и потребностей вероятных покупателей интернет-магазинов, являющиеся комфортным инструментом выбора при приобретении продуктов и предложений в онлайн-магазинах. Принципиально важными факторами, влияющими на развитие рекомендательных сервисов, являются польза и удобство одновременно и для потребителя, и для интернет-магазина. Пользователь, прежде всего, получает удобство интуитивного выбора. Для магазина открываются такие возможности, как увеличение среднего чека и выручки компании, альтернативная навигация во всем множестве товаров и получение информации о клиентах. Современные рекомендательные сервисы повышают наполненность онлайн-корзин на 12–60%, что обычно зависит от профильной направленности продукции.

**Ключевые слова:** коллаборативная фильтрация, гибридные системы, взвешенный наклон один (Weighted Slope One), модель Байеса, кластерная модель.

### Введение

Мы живем в мире информационного обилия. Выбор — это то, с чем встречается любой человек каждый день. При наличии всевозможных вероятностей в обязательном порядке появляется проблема выбора одной из альтернатив: какой продукт приобрести, какой кинофильм посмотреть, какое музыкальное произведение послушать, какую заметку прочитать и т. д. Рекомендательные системы — это комплект программных средств и способов, призванных посодействовать определенному пользователю в выборе между разными объектами [1].

Область использования рекомендательных систем многообразна. Наибольшее значение они имеют для интернет-магазинов. Персонализация онлайн-маркетинга считается отличительным трендом последних десятилетий [2]. По оценкам McKinsey, 35% дохода у Amazon и 75% у Netflix приходится как раз на рекомендуемые продукты, и данный процент, скорее всего, будет расти. Активно применяются рекомендательные системы и на веб-сайтах, производящих или публикующих контент [3].

Большинство клиентов интернет-магазинов, скорее всего, в будущем столкнется с использованием подобной программы, которая может дать нам личный совет и помочь в принятии решений [4]. Поиски новых методов создания подобной программы являются ключевой задачей той области науки, которая занимается разработкой рекомендательных систем. А задача самой системы — гарантировать сообществу пользователей предоставление доступной информации и качественных советов [5].

В связи с повсеместным внедрением онлайн и наличием большого объема статистических данных о предпочтениях пользователей научные работники всего мира стали активно заниматься разработкой этих алгоритмов в последние двадцать-тридцать лет [6]. А вследствие широкой применимости эти программы оказались достаточно эффективными.

В настоящей статье говорится о ведущих рекомендательных методах, а также приводятся примеры самых свежих разработок в данной области. Цель нашей работы — сделать обзор ведущих рекомендательных алгоритмов, рассказать о принципах способа

коллаборативной фильтрации, о системах фильтрации на базе содержания и на базе информации о пользователях, познакомить читателя с достоинствами и недостатками каждого из способов. Наиболее важным при знакомстве с данными разработками является понимание того, как их чаще всего используют, ведь все новейшие разработки в данной области направлены на улучшение именно этого параметра [7].

Для создания рекомендательных систем существует несколько методов, которые в основном базируются на коллаборативной фильтрации:

- Фильтрация, основанная на схожести пользователей;
- Фильтрация, основанная на схожести объектов (предметов);
- Фильтрация, основанная на модели;
- Модель Байеса;
- Регрессионная модель;
- Кластерная модель;
- Также достаточно распространены методы, основанные на факторизации;
- Неотрицательная матричная факторизация (NMF);
- Сингулярное разложение (SVD).

### Коллаборативная фильтрация

Коллаборативная фильтрация — это способ, позволяющий предугадывать неизвестные предпочтения пользователя на базе популярных оценок и/или поведения иных пользователей (Segaran 2007). Данный способ базируется на представлении о том, что пользователи, идентично оценившие одни предметы системы, имеют предрасположенность идентично оценивать и иные предметы системы. Еще одно допущение состоит в том, что пользователи дают схожие оценки предмету системы по парциальной шкале, к примеру, оценивая кинофильм от одной до десяти звезд на веб-сайте [imdb.com](http://imdb.com) (рис. 1). Данное допущение является важным, однако не во всех системах есть возможность собирать оценки пользователей. В этих случаях прибегают к скрытому сбору информации и оценке поведения, к примеру, записывая просмотренные ролики на YouTube и рекомендуя связанные с ними материалы [8].



**Dallas Buyers Club** (2013) 93

**8,0** Your rating: ★★★★★★☆☆ 9/10  
 Ratings: 8,0/10 from 146 639 users Metascore: 84/100  
 Reviews: 252 user | 389 critic | 47 from Metacritic.com

In 1985 Dallas, electrician and hustler Ron Woodroof works around the system to help AIDS patients get the medication they need after he is himself diagnosed with the disease.

**Director:** Jean-Marc Vallée  
**Writers:** Craig Borten, Melisa Wallack  
**Stars:** Matthew McConaughey, Jennifer Garner, Jared Leto  
[| See full cast and crew »](#)

[+ Watchlist](#) [Watch Trailer](#) [Share...](#)

**Won 3 Oscars.** Another 48 wins & 29 nominations. [See more awards »](#)

Рисунок 1 - Пример оценки фильма на сайте [imdb.com](http://imdb.com)

Коллаборативная фильтрация обычно делится на два подхода. Первым, и самым распространенным, считается подход, базирующийся на близости (сходстве) пользователей. Его сущность заключается в анализе прошлых оценок или поведения пользователя, исследование других пользователей, имеющих схожую “историю” и составление прогноза для неизвестных оценок [10].

В обычной ситуации образуется матрица «пользователи-предметы», смысл которой заключается в оценках определенным пользователем определенного предмета. Те ячейки, в которых нет значений, считаются неизвестными, то есть заявленному предмету пользователь не выставил оценку, значит, вероятнее всего, не воспользовался этим предметом (табл. 1).

Таблица – 1 Пример матрицы «пользователи-предметы»

	“Даллаский клуб покупателей”	“Человек-паук”	“Волк с Уолл-стрит”	“12 лет рабства”	“Free-to-play”
Пользователь_1n	5	3	5	-	-
Пользователь_2n	4	-	4	-	5
Пользователь_3n	-	1	5	2	-

Для определения близости пользователей используется некоторое количество различных алгоритмов, таких как:

- Манхэттенское расстояние;
- Евклидово расстояние;
- Коэффициент корреляции Пирсона.

Манхэттенское расстояние, или расстояние городских кварталов считается одним из базисных способов вычисления дистанции между 2-мя точками (1):

$$d(user1, user2) = \sum_{k=1}^n |user1_k - user2_k| \quad (1)$$

где

*user1, user2* – пользователи и их оценки;

*n* – количество предметов в матрице.

Данный подход является недостаточно точным при небольшом заполнении матрицы, но его преимущество в простоте и быстрой скорости выполнения.

Евклидово расстояние содержит квадратные корни (теорема Пифагора) и рассчитывается с помощью соответствующей формулы (2):

$$d(user1, user2) = \sqrt{\sum_{k=1}^n (user1_k - user2_k)^2} \quad (2)$$

где

*user1, user2* – пользователи и их оценки;

*n* – количество предметов в матрице.

Как и расстояние городских кварталов, этот способ содержит трудности при незаполненной матрице, но несложен в разработке и дешев в выполнении [10].

Более точный метод определения близости воплощен в коэффициенте корреляции Пирсона (3):

$$corr(user1, user2) = \frac{\sum_{i=0}^n (user1_i - \overline{user1})(user2_i - \overline{user2})}{\sqrt{\sum_{i=0}^n (user1_i - \overline{user1})^2} \sqrt{\sum_{i=0}^n (user2_i - \overline{user2})^2}} \quad (3)$$

где

$user1, user2$  – пользователи и их оценки;

$n$  – количество предметов в матрице.

$\overline{user1}$  и  $\overline{user2}$  – средняя оценка пользователей

Значение  $corr(user1, user2)$  может быть от -1 до 1, где -1 соответствует абсолютному несовпадению пользователей, а 1 — абсолютному совпадению.

Методы, базирующиеся на сходстве пользователей, могут быть полезны — они интуитивно понятны и просты в применении [10].

### Взвешенный наклон один (Weighted Slope One)

Одним из самых действенных алгоритмов в предметно-ориентированной коллаборативной фильтрации считается метод взвешенный наклон один (Weighted Slope One). Его сущность заключается в поиске различий оценок между парами элементов и применении данных различий для вычисления исключений [16]. Вычисление различий между предметами производится с помощью соответствующей формулы (4):

$$dev_{i,j} = \sum_{user \in S_{i,j}(X)} \frac{user_i - user_j}{both(S_{i,j}(X))} \quad (4)$$

где

$both(S_{i,j}(X))$  – число пользователей, оценивших и  $i$ -й, и  $j$ -й элемент;

$user$  – оценки пользователя;

Прогнозирование оценки предмета рассчитывается следующим образом (5):

$$P(user_j) = \frac{\sum_{i \in S(w-\{j\})} (dev_{j,i} + u_i) card(S_{j,i}(X))}{\sum_{i \in S(w-\{j\})} both(S_{j,i}(X))} \quad (5)$$

где

$both(S_{i,j}(X))$  – число пользователей, оценивших и  $i$ -й, и  $j$ -й элемент;

$user$  – оценки пользователя;

Взвешенный наклон один (Weighted Slope One) считается замечательным способом разработки рекомендательных систем. Имея невысокие требования к памяти и большую скорость работы, он демонстрирует высокую эффективность при наличии большого количества пользователей. Однако, сложной задачей для этого метода является проблема холодного старта, которая, в общем, относится ко всем предметно-ориентированным методам коллаборативной фильтрации [11]. Род алгоритмов Score One применяется в некоторых популярных сервисах, таких как hit flip, вебсайт рекомендаций DVD и Value Investing News, новостной портал фондовых бирж.

Второй большой группой методов коллаборативной фильтрации является фильтрация, базирующаяся на модели. Рассмотрим некоторые из них [11].

### Модель Байеса

Одним из самых популярных классификаторов является наивный байесовский классификатор. С его помощью создают рекомендации каких-то объектов. В его основе находится вероятностная модель теоремы Байеса [11]. Для работы этого алгоритма нужно сделать модель Байеса для каждого пользователя, который изучал какие-либо объекты, на основе содержания данных объектов (для кинокартин это могут быть артисты или жанры, для новостей — главные тексты и рубрики). Для нахождения более вероятной категории нужно определить относительные вероятности приспособления какого-нибудь предмета к любой категории и выбрать категорию, имеющую самую большую вероятность (6):

$$cat = \arg \max P(c) \prod P(o_i | c) \quad (6)$$

### Кластерная модель

Одним из самых популярных алгоритмов в кластерном анализе считается способ  $k$ -средних. Он используется при делении объектов или пользователей на группы — кластеры, формирующиеся по некоторым общим признакам, количество которых задается заранее. Сущность метода — в случайном выборе  $k$ -центров кластера и сокращении суммарного квадратичного отдаления пользователей или объектов от центра кластера [11]. Это рассчитывается с помощью соответствующей формулы (7):

$$d = \sum_i^k \sum_{x_j \in K} (x_j - u_i)^2 \quad (7)$$

где

$k$  – количество векторов;

$u$  – центр масс векторов из множества кластеров  $K$ .

### Факторизация матриц

Способы коллаборативной фильтрации довольно наглядны и просты, но с помощью факторизации матриц иногда можно получить гораздо больше результатов, например, данный способ позволяет обнаружить некоторые скрытые моменты, объединяющие объекты и пользователей. Но нужно помнить о том, что это математический способ, и для получения адекватных результатов потребуется настройка данного метода [11].

Целью неотрицательной факторизации матриц считается разложение матрицы на произведение 2-ух других матриц. В случае рекомендательной системы начальная матрица будет считаться матрицей «пользователи-объекты», а значения в ячейках — оценками данных пользователей всевозможных объектов. Если некоторых оценок нет, то с помощью факторизации вполне возможно получение данных недостающих оценок.

Математически это рассчитывается при помощи получения квадратичной погрешности, вычисления градиента от нее и получения значения матриц  $P$  и  $Q$ : (8)(9)

$$e_{i,j}^2 = (r_{i,j} - \sum_{k=1}^K p_{i,j} q_{i,j})^2 \quad (8)$$

где

$r$  – реальное значение исходной матрицы,

$p$  и  $q$  – значения предполагаемых матриц  $P$  и  $Q$

$$p_{i,k}^{new} = p_{i,k} + 2\alpha e_{i,j} q_{k,j} \quad (9)$$

$$q_{k,j}^{new} = q_{k,j} + 2\alpha e_{i,j} p_{i,k}$$

где

$p_{i,k}^{new}$  и  $q_{k,j}^{new}$  – новые значения в матрицах  $P$  и  $Q$

$2\alpha e$  – градиент из квадратичной ошибки

Этот метод требует большого числа итераций (>5000) для получения верных итогов.

### Проблемы существующих подходов

#### • Недостаток информации о пользователях

Основная масса пользователей различных сервисов не любит высказывать свое мнение о каких-либо предметах и выставлять им оценки, из-за чего точность прогноза утрачивается.

#### • Новые пользователи

Данная проблема также является серьезной — новый пользователь не предоставляет достаточно информации для создания качественных рекомендаций для него, поэтому рекомендации формируются или из известного в целом контента, или не формируются вообще.

• *Поддельные оценки*

На многих популярных порталах есть определенные пользователи, которые осознанно завышают или занижают оценки всем объектам системы, продвигая те или иные продукты, что мешает работать рекомендационным алгоритмам.

• *Масштабируемость*

При увеличении количества пользователей и объектов системы возрастает время выполнения алгоритмов для каждого пользователя. Отчасти это решается улучшением алгоритмов (например, для метода *weighted slope one* нужно только наблюдать за отклонениями между парами элементов и общим количеством элементов).

В итоге можно сказать, что все существующие методы имеют как плюсы, так и минусы. Одни методы выигрывают по временным затратам, другие — по требованиям к памяти и точности, но для данной работы было принято решение выбрать способы, базирующиеся на статистических вычислениях. Создание классификатора не всегда приводит к нужным результатам. Например, важные тексты не всегда попадают в анонсы, а ведущие новостей не дают теги к объектам новости. Оценивание каждого новостного контента, прочтенного пользователем, не всегда дает верное заключение о его предпочтениях.

Поэтому для создания рекомендательной системы было принято решение выбрать метод поиска схожих людей через взвешенный наклон один (*Weighted Slope One*).

**Постановка задачи**

Сформулируем задачу. Допустим, у нас будет большое количество пользователей  $U = \{u_1, u_2, \dots, u_n\}$ , большое количество объектов  $P = \{p_1, p_2, \dots, p_m\}$  и матрица рейтингов  $R = (r_{i,j})$  размера  $n \times m$ , где  $i \in 1 \dots n, j \in 1 \dots m$ . Вероятными значениями рейтингов могут быть числа от 1 (не нравится) до 5 (нравится). Если пользователь  $i$  не оценил элемент  $j$ , то на месте  $r_{i,j}$  значение будет пустующим. Описанную матрицу рейтингов можно представить в виде таблицы (табл. 2). Обозначим через  $r'_{i,j}$  наш прогноз относительно того, какую оценку пользователь  $i$  поставит продукту  $j$ . Наша задача предсказать, какие оценки  $r_{i,j}$  должны стоять на месте пропусков в матрице, то есть рассчитать  $r'_{i,j}$ .

Таблица – 2 Пример матрицы «пользователи-продукты»

Продукты	Продукт 1	Продукт 2	Продукт 3	Продукт 4	Продукт 5	Продукт 6
Пользователь 1	2	4	5	3	4	?
Пользователь 2	1	?	1	4	5	?
Пользователь 3	3	4	?	1	4	?
Пользователь 4	?	?	4	2	?	2
Пользователь 5	?	4	5	3	?	?

Затем для каждого пользователя  $u$  на основе спрогнозированных оценок  $r'_{i,j}$ , нам нужно сформировать список из  $N$  продуктов, которые соответствуют предпочтениям пользователя и которые еще им не оценены. Список этих  $N$  продуктов обозначим через  $N$ -мерный вектор  $(p_{i_1}, p_{i_2}, \dots, p_{i_n})$ . Таким образом, математическая задача звучит так:

**Дано:**

$U = \{u_1, u_2, \dots, u_n\}$  — множество пользователей,

$P = \{p_1, p_2, \dots, p_m\}$  — множество продуктов,

$R = (r_{i,j})$  - матрица рейтингов размера  $n \times m$ , где на месте  $r_{i,j}$  будет стоять число, если пользователь  $u_i$  оценил продукт  $p_j$  и пусто в ином случае.  $N$  — требуемое число рекомендаций, которые нужно получить от системы.

Требуется найти: Для данного пользователя  $u_i$  найти  $N$ -мерный вектор  $(p_{i_1}, p_{i_2}, \dots, p_{i_n})$ , где продукты  $p_{i_k}$ ,  $k \in N$  еще не оценены этим пользователем, то есть в матрице рейтингов  $R = (r_{i,j})$  пусто на месте  $r'_{i,i_k}$ , а также проследить, чтобы эти продукты наиболее точно удовлетворяли предпочтения пользователя, то есть прогнозные рейтинги  $r'_{i,i_k}$  были наибольшими.

### Решение:

Взвешенный наклон один (Weighted Slope One)

Вычисление различий между продуктами производится с помощью соответствующей формулы (10):

$$dev_{i_k,j} = \frac{1}{|U_{i_k,j}|} \sum_{u \in \{U_{i_k,j}\}} (r_{u,i_k} - r_{u,j}) \quad (10)$$

где

$|U_{i_k,j}|$  — число пользователей, оценивших и  $i_k$ -й, и  $j$ -й элемент;

$\{U_{i_k,j}\}$  — набор пользователей, оценивших и  $i_k$ -й, и  $j$ -й элемент;

$u$  — оценки пользователя.

Прогноз оценки предмета рассчитывается следующим образом (11):

$$r'_{i,i_k} = \frac{1}{|S_u - \{i_k\}|} \sum_{j \in \{S_u - \{i_k\}\}} (dev_{i_k,j} + r_{i,j}) \quad (11)$$

где

$\{S_u - \{i_k\}\}$  — все элементы, оцененные пользователем  $u$ , кроме элемента  $i_k$

$|S_u - \{i_k\}|$  — представляет количество элементов, оцененных пользователем  $u$ , кроме элемента  $i_k$

$r_{i,j}$  — рейтинг пользователя  $u$  для элемента  $j$ .

Итак, спрогнозируем не оцененные товары пользователем  $u_i$ .

После создания прогноза необходимо составить список Top N ( $N$ -мерный вектор) для рекомендаций.

Итого, есть все прогнозы для всех продуктов, которые пользователь  $u_i$  еще не оценил.

Теперь составим список Top N ( $N$ -мерный вектор):

1. Пусть рейтинги будут от 1 до 5.
2. Если прогноз больше 5, мы посчитаем, что это 5.
3. Нужен лимит рекомендаций, поставим 3.5.

После этого добавим в список Top N ( $N$ -мерный вектор) все продукты с прогнозом от 3,5 до 5 и будем рекомендовать этот список с продуктами для пользователя  $u_i$ .

### Заключение

В настоящей работе рассмотрены главные разновидности рекомендательных систем и основы их создания. Приведен детальный перечень алгоритмов коллаборативной фильтрации, показаны методы оценки свойств схожих объектов.

В качестве основного способа решения задачи выбран способ, основанный на коллаборативной фильтрации и реализованы методы, базирующиеся как на близости пользователей и объектов, так и на факторизации исходных данных. Наилучший результат продемонстрировал метод взвешенный наклон один (Weighted Slope One).

Работа может быть продолжена путем совершенствования базисных алгоритмов, проведения экспериментов с построением других гибридных моделей, применения

добавочных метаданных для создания систем, основанных на информации о пользователе, и систем на базе контента.

#### СПИСОК ЛИТЕРАТУРЫ

1. М.А.Терехин, Разработка системы рекомендаций. [Электронный ресурс] URL: <https://elib.bsu.by/bitstream/123456789/232575/1/118-121.pdf> (accessed: 2019)
2. Mackenzie I., Meyer C., Noble S. How retailers can keep up with consumers. [Электронный ресурс] URL: <https://www.mckinsey.com/industries/retail/our-insights/howretailers-can-keep-up-with-consumers> (дата обращения: 26.03.2021)
3. Daniel Lemire, Anna Maclachlan. «Slope One Predictors for Online Rating-Based Collaborative Filtering.» In SIAM Data Mining (SDM'15), 2019. [Электронный ресурс] URL: [https://www.researchgate.net/publication/326079280\\_A\\_trust-based\\_collaborative\\_filtering\\_algorithm\\_for\\_E-commerce\\_recommendation\\_system](https://www.researchgate.net/publication/326079280_A_trust-based_collaborative_filtering_algorithm_for_E-commerce_recommendation_system) (дата обращения: 26.03.2021)
4. Lefats'e Manamolela, Tranos Zuva, Martin Appiah. Collaborative Filtering Recommendation Systems Algorithms, Strengths and Open Issues, 2020. [Электронный ресурс] URL: [https://link.springer.com/chapter/10.1007%2F978-3-030-63319-6\\_14](https://link.springer.com/chapter/10.1007%2F978-3-030-63319-6_14) (дата обращения: 26.03.2021)
5. Pawel Matuszek, João Vinagre, Myra Spiliopoulos «Forgetting methods for incremental matrix factorization in recommender systems.» Conference: ACM SAC 2015, [Электронный ресурс] URL: [https://www.researchgate.net/publication/280246432\\_Forgetting\\_methods\\_for\\_incremental\\_matrix\\_factorization\\_in\\_recommender\\_systems](https://www.researchgate.net/publication/280246432_Forgetting_methods_for_incremental_matrix_factorization_in_recommender_systems) (дата обращения: 26.03.2021)
6. Sanderson, Dan. Programming Google App Engine. O'Reilly Media, 2019. [Электронный ресурс] URL: <https://pdfweek.com/downloads/programming%20google%20app%20engine%20programming%20gamenetore%20pdf> (дата обращения: 26.03.2021)
7. Segaran, Toby. Programming Collective Intelligence. O'Reilly Media, 2017. O'Reilly Media, Inc, – Т. 13, No. 1. – P.117–134.
8. Hill W., Stead L., Rosenstein M., Furnas G. Recommending and Evaluating Choices in a Virtual Community of Use // Proceeding Conference Human Factors in Computing Systems, 2015. P. 194–201.
9. Resnick P., Iakovou N., Sushak M., Bergstrom P., Riedl J. GroupLens: An Open Architecture for Collaborative Filtering of Netnews // Proceeding 1994 Computer Supported Cooperative Work Conference, 2014. P. 175–186.

#### Найзабаева Л., Алашыбаев Б.А.

#### Машиналық оқытуды қолдану арқылы интернет-дүкендерге арналған ұсыныс жүйесі

**Аңдатпа.** Интернет-маркетингтің соңғы тенденцияларының арасында ұсыныс жүйелерін бөлуге болады. Ұсынушы жүйелер – бұл интернет-дүкендердің тауарлары мен қызметтерін сатып алу кезінде таңдаудың ыңғайлы құралы болып табылатын интернет-дүкендердің әлеуетті клиенттерінің қызығушылықтары мен қажеттіліктерін болжауға бағытталған арнайы қосымшалар. Ұсыныс қызметтері пайдаланушы үшін де, интернет-дүкен үшін де пайдалы және ыңғайлы болуы өте маңызды. Ең алдымен қолданушының ыңғайлы және интуитивті таңдауы бар. Сонымен қатар дүкенге бару кезінде орташа чек пен кірісті ұлғайту, тауарлардың алуан түріндегі альтернативті навигация және тұтынушылар туралы дереккөзі ақпарат мүмкіндіктерін ашады. Бұл, әдетте өнімнің профиліне байланысты, бүгінгі таңда заманауи ұсынымдық қызметтер онлайн-дүкен арбаларының құрамын 12-60%-ға арттырады.

**Түйінді сөздер:** бірлескен сүзу, гибридті жүйелер, өлшенген көлбеу, Байес моделі, кластерлік модель.

**Naizabayeva L., Alashybayev B.A.**

**A recommendation system for online stores using Machine Learning**

**Abstract.** Recommender systems can be singled out among the latest trends in Internet marketing. Recommender systems are special applications focused on predicting the interests and needs of potential customers of online stores, which are convenient tools for choosing when buying goods and services in online stores. It is fundamentally important that recommendation services are useful and convenient for both the user and the online store. The user, first of all, has the convenience and intuitiveness of the choice. At the same time, the store opens up such opportunities as increasing the average check and revenue per visit, alternative navigation in the entire variety of products and a source of customer information. Today, modern recommendation services increase the content of online shopping carts by 12-60%, which usually depends on the profile of the product.

**Keywords:** collaborative filtration, hybrid systems, Weighted Slope One, Bayesian model, Cluster model.

**Авторлар туралы мәлімет:**

**Найзабаева Лязат**, т.ғ.д., Халықаралық ақпараттық технологиялар университеті, «Ақпараттық жүйелер» кафедрасының қауымдастырылған профессоры.

**Алашыбаев Бекарыс Алашыбаевич**, «Ақпараттық жүйелер» кафедрасының магистранты, Халықаралық ақпараттық технологиялар университеті.

**Сведения об авторах:**

**Найзабаева Лязат**, доктор технических наук, ассоциированный профессор кафедры «Информационные системы», Международный университет информационных технологий.

**Алашыбаев Бекарыс Алашыбаевич**, магистрант кафедры «Информационные системы», Международный университет информационных технологий.

**About the authors:**

**Lyazat Nayzabayeva**, Dr. Sc. (Technology), Associate Professor, Department of Information Systems, International Information Technology University.

**Bekarys A. Alashybayev**, master student, Department of Information Systems, International Information Technology University.

INTERNATIONAL JOURNAL OF INFORMATION AND  
COMMUNICATION TECHNOLOGIES

МЕЖДУНАРОДНЫЙ ЖУРНАЛ ИНФОРМАЦИОННЫХ И  
КОММУНИКАЦИОННЫХ ТЕХНОЛОГИЙ

ХАЛЫҚАРАЛЫҚ АҚПАРАТТЫҚ ЖӘНЕ  
КОММУНИКАЦИЯЛЫҚ ТЕХНОЛОГИЯЛАР ЖУРНАЛЫ

Ответственный за выпуск	Есбергенов Досым Бектенович
Редакторы	Далабаева Айсара Касымбековна Джоламанова Балия Джалгасбаевна Медведев Евгений Юрьевич
Компьютерная верстка	Туратауова Айжаркын Ахметовна
Компьютерный дизайн	Туратауова Айжаркын Ахметовна

Редакция журнала не несет ответственности за  
недостоверные сведения в статье и  
неточную информацию по цитируемой литературе

Подписано в печать 26.06.2021 г.  
Тираж 500 экз. Формат 60x84 1/16. Бумага тип.  
Уч.-изд.л. 10.1. Заказ №165

Издание Международный университет информационных технологий  
Издательский центр КБТУ, Алматы, ул. Толе би, 59