

О.Ю. Чередніченко, О.В. Янголенко, О.В. Іващенко, О.М. Матвеев

Національний технічний університет “Харківський політехнічний інститут”, Харків

МОДЕЛІ ФОРМУВАННЯ РЕКОМЕНДАЦІЙ У ІНТЕЛЕКТУАЛЬНИХ СИСТЕМАХ ЕЛЕКТРОННОЇ КОМЕРЦІЇ

Результати роботи пошукових та фільтраційних механізмів сучасних систем електронної комерції не завжди задовольняють вимоги користувачів, що проявляється у неточних та неповних рекомендаціях товарів за пошуковим запитом. Удосконалення якості рекомендацій для покупок онлайн торговельних платформ є актуальною задачею. Дана робота наводить моделі формування рекомендацій на основі методів кластерного аналізу, які дозволяють згрупувати схожі товари та схожих клієнтів за їхніми характеристиками. Наведені результати експерименту щодо формування рекомендацій придбання рюкзаків для покупок онлайн магазину спортивного обладнання.

Ключові слова: E-commerce система, рекомендаційна інтелектуальна система, кластеризація, міри відстані.

Вступ

На фоні постійного зростання об'ємів інформації, доступної у мережі Інтернет, користувачам стає все важче працювати з нею. Це стосується, зокрема, і систем електронної комерції (e-commerce systems), де знаходять місце мільйони пропозицій різноманітних товарів. Тому велику роль почали відігравати системи, які допомагають користувачам орієнтуватися у великому просторі інформації. Людина не має можливості аналізувати велику кількість інформації, адже це складно та потребує багато часу і фізичних зусиль. Але завдяки рекомендаційним системам, які здатні фільтрувати велику кількість інформації, та надавати користувачу саме потрібну йому інформацію та рекомендації за його вподобаннями, цю проблему можна вирішити.

Рекомендаційні системи вже досить поширені і мають велику кількість застосувань. В першу чергу, рекомендаційні системи використовуються в інтернет-комерції для того, щоб допомогти користувачам вибрати відповідні товари. Такі сервіси збирають інформацію про вподобання користувачів і намагаються запропонувати їм корисні товари. Яскравими прикладами компаній, що використовують даний підхід, є Amazon, eBay, iTunes та інші. Інша застосування рекомендаційних систем – це допомога користувачам у виборі книг, музики і фільмів. Наприклад, сервіси Pandora, GoodReads, and IMDb використовують рекомендаційні системи для цих цілей. Рекомендаційні системи – клас інформаційно-пошукових систем, які дають змогу передбачити, які об'єкти будуть цікавими та корисними користувачу (надати користувачу рекомендації), відштовхуючись від конкретної інформації про профіль користувача.

Рекомендаційні системи змінили способи взаємодії веб-сайтів зі своїми користувачами. Замість надання статичної інформації, коли користувачі шу-

кають і, можливо, купують продукти, рекомендаційні системи збільшують ступінь інтерактивності для розширення послуг користувачеві можливостей. Рекомендаційні системи формують рекомендації незалежно для кожного конкретного користувача на основі його минулих оцінок і пошуків, а також на основі поведінки інших користувачів [1]. Тому, зараз такі системи стали невід'ємною частиною майже будь-якого веб-сайту.

Отже, з розвитком та активним використанням сучасних інтернет-технологій, зокрема платформ електронної комерції, актуальними стали веб-орієнтовані рекомендаційні системи.

Проте не зважаючи на достатнє поширення існуючих рекомендаційних систем в електронній комерції, якість їхньої роботи викликає питання, що може відчувати на собі будь-який користувач торговельних онлайн платформ. Товари, що пропонуються внаслідок фільтрації та відповідно до вподобань користувача, часто не є такими, що відповідають усім його вимогам. Частина товарів, яка зареєстрована у e-commerce системі, взагалі не потрапляє до результатів пошуку. Через це користувач позбавляється можливості обирати із повного переліку наявних товарів.

Таким чином, задача вдосконалення моделей формування рекомендацій у системах електронної комерції є актуальною та вимагає нових підходів до її вирішення.

Аналіз досліджень в області моделювання рекомендаційних систем. На сьогодні існує велика кількість робіт, присвячених розробці алгоритмів та методів формування рекомендацій для систем електронної комерції.

У роботах [2–3] розглянуто задачу моделювання результатів пошуку рекомендаційної системи на основі критерія ціни товару та враховуючи цінову

конкуренцію. Питання сегментації електронного ринку з точки зору покупців та продавців досліджуються у роботах [4–5]. Автори побудували модель, що дозволяє збільшити об'єми транзакцій.

Задля підвищення точності рекомендацій багатьох авторів звертається до задачі структуризації опису товарів на онлайн платформах. Наприклад, у роботах [6–7] розглядаються питання нормалізації атрибутів товарів та видалення дублікатів.

Проблеми недосконалості існуючих пошукових алгоритмів, що лежать в основі e-commerce систем, розглянуто в роботах [7–8]. Зокрема, основну увагу приділено питанням кластеризації товарів відповідно до їхнього опису на групи подібних з метою надавати користувачу можливість обрати товар з найкращими умовами продажу. Автори дійшли висновку, що задля підвищення точності кластеризації необхідно здійснювати попередню обробку

Не зважаючи на те, що питанням розробки рекомендаційних систем присвячено достатньо робіт, проте задача удосконалення якості рекомендацій лишається відкритою. Тому **метою даної роботи** є підвищення якості результатів пошуку товарів в системах електронної комерції за рахунок використання інтелектуального підходу до кластеризації користувачів відповідно до їхніх вподобань та самих товарів відповідно до їхньої схожості за власними атрибутами.

Виклад основного матеріалу

Моделі формування рекомендацій

Задача формування рекомендацій на онлайн торговельній платформі розбивається на декілька етапів:

1. Парсинг контенту сайту електронного магазину та збереження даних щодо товарів у JSON файлах.
2. Формування паралельного корпусу, тобто набору даних на мові оригіналу та у перекладеному вигляді.
3. Збір атрибутів товарів та їхніх значень.
4. Генерація тезауруса.
5. Нормалізація зібраних даних.
6. Генерація рекомендацій для користувача.

У даній роботі пропонується використовувати інструментарій кластерного аналізу для формування користувацьких рекомендацій. Застосування кластерного аналізу в загальному виді зводиться до наступних етапів [10].

1. Відбір вибірки об'єктів для кластеризації.
2. Визначення безлічі змінних, за якими будуть оцінюватися об'єкти у вибірці. При необхідності – нормалізація значень змінних.
3. Обчислення значень міри схожості між об'єктами.

4. Застосування методу кластерного аналізу для створення груп схожих об'єктів (кластерів).

Для того, щоб визначити “схожість” об'єктів, для початку потрібно скласти вектор характеристик для кожного об'єкта – як правило, це набір числових значень, наприклад, ріст-вага людини. Однак існують також алгоритми, що працюють з якісними (категорійними) характеристиками. Для кожної пари об'єктів вимірюється “відстань” між ними – ступінь схожості, наприклад, Евклідова відстань, Манхеттенська відстань, відстань Чебишева і т.д. В даній роботі використовуються Евклідова відстань [10].

Згідно Евклідової відстані схожість або відмінність між об'єктами встановлюється в залежності від метричної відстані між ними. Всі змінні стандартизовані, не мають розмірності. Це забезпечує можливість підсумовування відстаней за різними змінним. Іншими словами, це геометрична відстань між двома об'єктами в багатовимірному просторі. Обчислення багатовимірної Евклідової відстані частіше називають простою Евклідовою відстанню.

Повністю алгоритм формування рекомендації представлено на рис. 1. На першому етапі визначається множина характеристик товарів:

$$x_{ij}^k (i \in I; j \in J; k \in C), \quad (1)$$

де I – множина товарів;

J – множина характеристик товару;

C – множина клієнтів;

$x_{ij}^k (i \in I; j \in J; k \in C)$ – оцінка j -ї характеристики

ки i -го товару k -м клієнтом;

$$i = \overline{1, a}; j = \overline{1, z}; k = \overline{1, n}.$$

Другий етап – складається з двох частин, спочатку нормалізуються тільки порядкові та кількісні (неперервні) вхідні дані (оцінки):

$$x_{ij}^c = \frac{x_{ij}^c}{\max_{c \in C} \{x_{ij}^c\}}, \quad (2)$$

де x_{ij}^c – нормалізовані оцінки всіх клієнтів c ($c \in C$)

за j -ю характеристикою i -го товару ($j \in J, i \in I$).

Друга частина другого етапу – визначаємо відстань d_i^{km} за Евклідовою метрикою між клієнтами k та m з урахуванням коефіцієнта важливості w_j^k ($0 \leq w_j^k \leq 1$):

$$d_i^{km} = \sqrt{\sum_{j=1}^n (w_j^k (x_{ij}^k - x_{ij}^m))^2}, \quad (3)$$

де x_{ij}^k, x_{ij}^m ($k, m \in C$) – нормалізовані оцінки всіх клієнтів $k, m \in C$ за j -ю характеристикою i -го товару ($j \in J, i \in I$);

– коефіцієнт важливості w_j^k ($0 \leq w_j^k \leq 1$).

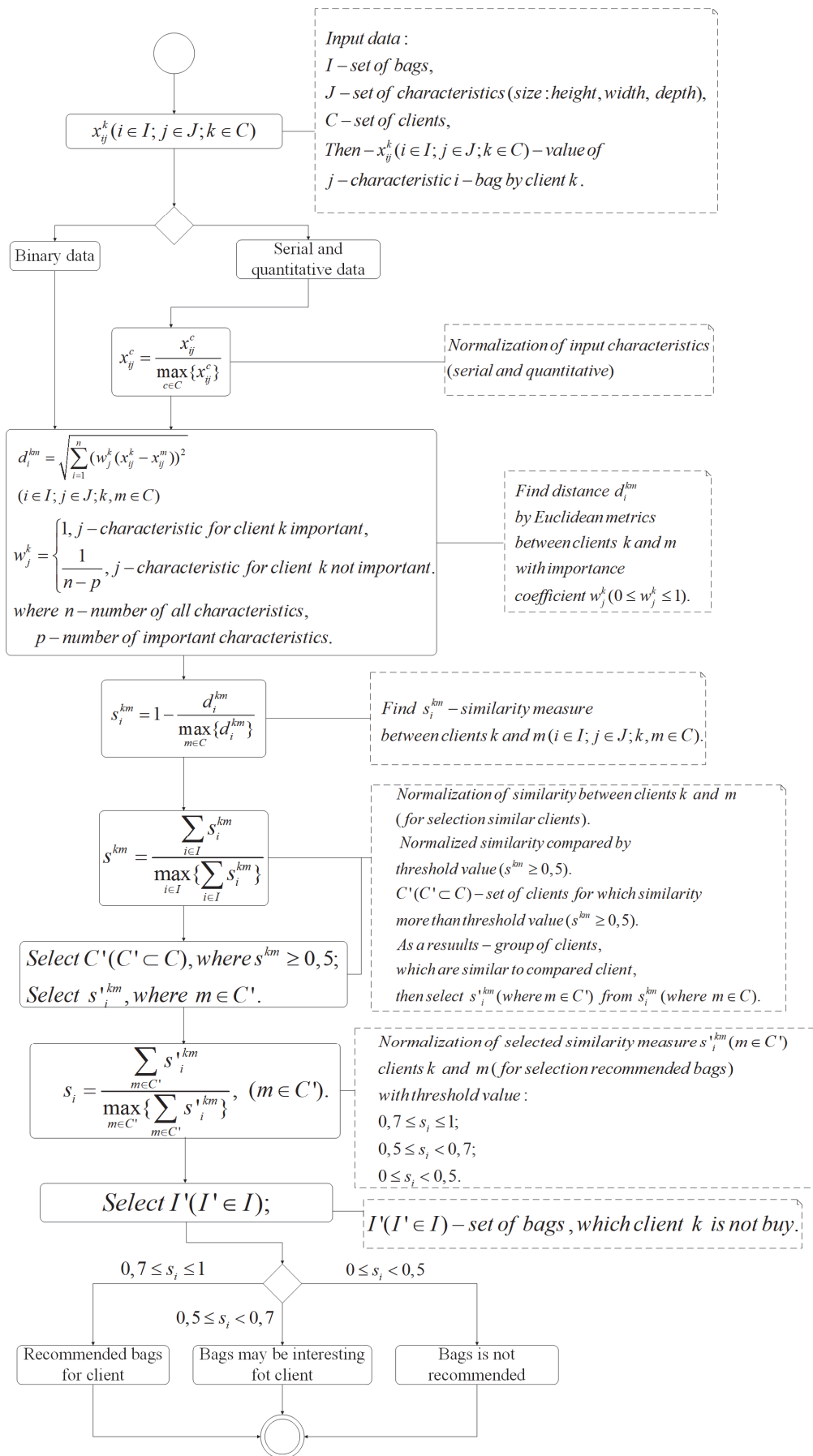


Рис. 1. Алгоритм формування рекомендації

Третій етап – визначення міри подібності s_i^{km} клієнтів k та m ($i \in I; j \in J; k, m \in C$):

$$s_i^{km} = 1 - \frac{d_i^{km}}{\max_{m \in C} \{d_i^{km}\}}, \quad (4)$$

де d_i^{km} – відстань за Евклідовою метрикою між клієнтами k та m з урахуванням коефіцієнта важливості w_j^k ($0 \leq w_j^k \leq 1$) (з другого етапу).

Четвертий етап – нормалізація отриманої міри подібності (s_i^{km}) (отриманої у третьому етапі) для відбору схожих клієнтів:

$$s^{km} = \frac{\sum_{i \in I} s_i^{km}}{\max_{i \in I} \{ \sum_{i \in I} s_i^{km} \}}, \quad (5)$$

де s^{km} – нормалізація отриманої міри подібності клієнтів k та m (за Евклідовою метрикою);

s_i^{km} – міра подібності клієнтів k та m (за Евклідовою метрикою), отриманої у третьому етапі.

П'ятий етап – відбір схожих клієнтів (щодо клієнта, який порівнюється), отримані результати у четвертому етапі порівнюються із заданим пороговим значенням ($s^{km} \geq 0,5$), тобто відбираємо множину клієнтів, для яких міра подібності більша за порогове значення, результатом даного порівняння є група клієнтів, які схожі з клієнтом, що розглядається – тобто відбираються s_i^{km} (де $m \in C'$) із s_i^{km} (де $m \in C$), але тільки для тих клієнтів, які пройшли відбір за пороговим значенням.

Шостий етап – нормалізація відібраної отриманої міри подібності ($s_i^{'km}$) (отриманої у п'ятому

етапі) для відбору рекомендованих товарів із заданими пороговими значеннями ($0,7 \leq s_i \leq 1$; $0,5 \leq s_i < 0,7$; $0 \leq s_i < 0,5$):

$$s_i = \frac{\sum_{m \in C'} s_i^{'km}}{\max_{m \in C'} \{ \sum_{m \in C'} s_i^{'km} \}}, (m \in C'), \quad (6)$$

де s_i – нормалізація отриманої міри подібності клієнтів k та m (за Евклідовою метрикою);

$s_i^{'km}$ (де $m \in C'$) – відібрана за пороговим значення міра подібності клієнтів k та m (за Евклідовою метрикою) отриманої у п'ятому етапі.

Сьомий (останній) етап – відбір товарів $I'(I' \in I)$ – множина товарів, які клієнт, що розглядається (k) не відвідував – відбір із шостого етапу та створення із них рекомендації із заданими пороговими значеннями ($0,7 \leq s_i \leq 1$) – рекомендувати товари клієнту; $0,5 \leq s_i < 0,7$ – товари можуть бути цікавими клієнту; $0 \leq s_i < 0,5$ – не рекомендувати товари клієнту).

Результати експерименту

В якості експерименту в даній статті аналізуються дані онлайн торгівельної платформи продажу спортивних товарів. Вхідні дані для експерименту сформовані із множини рюкзаків, дані щодо яких були збережені у JSON форматі. Всього товарів на момент проведення експерименту – 141, мова – німецька. Вхідні дані, що були нормалізовані, виглядають, як показано на рис. 2. До атрибутів, які було виділено належать: Brand, Price, Technology_Material, Equipments, Other, LoadRange, Size, Volume, Weight, BackSystem, Function, Material, Note, Filling, WorkingPressure, Volume_Weight.

```
[
  {
    "Brand": ["Kohla Zugspitze 26", "Shimano Unzen Enduro 4 1"],
    "Price": ["€ 69.99", "€ 94.99"],
    "Technology_Material": ["Surround Ventilation System", "Rider Fit X-Harness, Shaped Air Mesh Back Part"],
    "Other": ["Stockhalterung", "Hydration Bladder: incl. 2l-Hydration Bladder"],
    "LoadRange": ["0 - 4 kg", "3 - 8 kg"],
    "Volume": ["9.0 l", "28 liters"],
    "Weight": ["740 g", "380 g"],
    "BackSystem": ["Ventilation Foam Back Optimizes Deduction of Moisture", "A.I.R. Back Ventilation System"],
    "Function": ["hydration system compatible", "superlight, functional"],
    "Equipments": [
      "removable compression straps, stowable attachment loops for ice picks or trekking poles",
      "including rain cover, Shove-It bag for quick stowage and removal of equipment, attachment loops"],
    "Material": ["Dynajin 210, 30% Polyester / 70% Polyamide", "Body: 95% Polyester, 5% Nylon, Lining: 100% Polyester"],
    "Note": [
      "Cartridge is not included",
      "Shipping only to a delivery address in Austria or a Hervis branch (Click & Collect) possible"],
    "Filling": ["nitrogen (only factory filling possible)"],
    "WorkingPressure": ["300 bar"],
    "Size": ["75 x 36 x 30 cm", "65 x 20 x 30 cm"],
    "Volume_Weight": "10L + 3kg"
  }
]
```

Рис. 2. Нормалізовані вхідні дані щодо товарів

Результати етапів виконання алгоритму формування рекомендації наведені в табл. 1–8. Таким чином, до рекомендованих товарів належить ті, що

були відібрані у сьомому, восьмому та десятому турах.

Таблиця 1

Очікувані результати етапу 1 (вхідні дані)

№ туру	Характеристика	№ клієнта						
		1	2	3	4	5	6	7
1	Ціна (Price)	4	5	3	5	4	0	0
	Вага (Weight)	1	1	1	1	1	0	0
	Об'єм (Volume)	1	2	3	2	3	0	0
	Розмір (Size)	0	0	1	0	1	0	0
	Бренд (Brand)	3	2	4	5	4	0	0

Таблиця 2

Очікувані результати етапу 2 (нормалізація вхідних даних)

№ туру	Характеристика	№ клієнта						
		1	2	3	4	5	6	7
1	Ціна (Price)	0,8	1	0,6	1	0,8	0	0
	Вага (Weight)	1	1	1	1	1	0	0
	Об'єм (Volume)	0,33	0,67	1	0,67	1	0	0
	Розмір (Size)	0	0	1	0	1	0	0
	Бренд (Brand)	0,6	0,4	0,8	1	0,8	0	0

Таблиця 3

Очікувані результати етапу 3 (розрахунок Евклідової відстані)

№ туру	№ порівняння 1-го клієнта із:					
	2	3	4	5	6	7
1	0,49	1,6	0,56	1,22	1,94	1,94
2	1,9	1,9	1,45	1,26	1,9	1,9
3	2,47	0	1,48	0	1,48	1,83
4	1,11	1,36	1,1	1,59	2,4	2,4
5	2,3	2,65	2,45	2,45	2,45	2,45
6	0,57	0,28	0,4	1,48	0,7	0,49
7	0	0	2,45	0	0	2,29
8	0	0	0	2,65	1,87	2,65
9	0	0,75	0,39	0,2	2,65	2,65
10	0	2,3	0	0	2,65	0

Таблиця 4

Очікувані результати етапу 4 (розрахунок міри подібності)

№ туру	№ порівняння 1-го клієнта із:					
	2	3	4	5	6	7
1	0,75	0,18	0,71	0,37	0	0
2	0	0	0,24	0,34	0	0
3	0	1	0,4	1	0,4	0,26
4	0,54	0,43	0,54	0,34	0	0
5	0,13	0	0,08	0,08	0,08	0,08
6	0,61	0,81	0,73	0	0,53	0,67
7	1	1	0	1	1	0,07
8	1	1	1	0	0,29	0
9	1	0,72	0,85	0,92	0	0
10	1	0,13	1	1	0	1

Таблиця 5

Очікувані результати етапу 5 (нормалізація даних – клієнтів)

№ клієнта					
2	3	4	5	6	7
1	0,87	0,92	0,84	0,38	0,34

Таблиця 6

Очікувані результати етапу 6 (результат після нормалізації даних клієнтів)

№ туру	№ порівняння 1-го клієнта із:			
	2	3	4	5
1	0,75	0,18	0,71	0,37
2	0	0	0,24	0,34
3	0	1	0,4	1
4	0,54	0,43	0,54	0,34
5	0,13	0	0,08	0,08
6	0,61	0,81	0,73	0
7	1	1	0	1
8	1	1	1	0
9	1	0,72	0,85	0,92
10	1	0,13	1	1

Таблиця 7

Очікувані результати етапу 7 (результат нормалізації даних товарів)

№ туру/результат нормалізації									
1	2	3	4	5	6	7	8	9	10
0,58	0,17	0,69	0,53	0,08	0,62	0,86	0,86	1	0,9

Таблиця 8

Очікувані результати етапу 8 (рекомендації)

Категорія рекомендації	№ рекомендованих турів
Рекомендовані	7, 8, 10
Можуть бути цікавими	3
Не рекомендовані	5

Висновки

Формування якісних рекомендацій для покупців онлайн торговельних платформ є результатом реалізації ефективних пошукових та фільтраційних алгоритмів систем електронної комерції. Існуючі методи не завжди надають клієнтам повні списки товарів, що відповідають їхнім вимогам. У даній роботі запропоновано підхід до формування рекомендацій, заснований на кластерному аналізі товарів та клієнтів. Представлені моделі дозволяють, з одного боку, знайти групи товарів, схожих за значен-

нями атрибутів, та з іншого боку, сформувати класи клієнтів, що мають схожі потреби.

Перевагою запропонованого підходу є те, що на вхід алгоритму кластеризації подаються описи товарів, видобуті з сайту електронного магазину, що були попередньо нормалізовані. Нормалізація товарів передбачає приведення їхніх атрибутів та значень атрибутів до стандартизованого вигляду.

Це дозволяє надалі підвищити точність та повноту пошукового алгоритму та якість рекомендацій в цілому.

Таким чином, розроблені інтелектуальні моделі та алгоритми дозволили реалізувати процес формування рекомендацій, що було продемонстровано в експерименті із вибором товарів онлайн магазину спортивного обладнання. У подальшому планується запропоновані моделі використати для пошуку товарів-дублікатів, що штучно створюються продавцями систем електронної комерції задля збільшення своєї присутності на електронному ринку.

Список літератури

1. Introduction to Electronic Commerce and Social Commerce / E. Turban, J. Whiteside, D. King, J. Outland. – Springer. – 2017. – 413 p. <https://doi.org/10.1007/978-3-319-50091-1>.
2. Consumer Price Search and Platform Design in Internet Commerce / M. Dinerstein, L. Einav, J. Levin, N. Sundaresan // American Economic Review. – 2018. – № 108. – P. 1820-1859. <https://doi.org/10.1257/aer.20171218>.

3. Hortacsu A. Testing Models of Consumer Search using Data on Web Browsing and Purchasing Behavior / A. Hortacsu, De los S. Babur, W. Matthijs // *American Economic Review*. – 2012. – № 102(6). – P. 2955-2980. <https://doi.org/10.1257/aer.102.6.2955>.
4. Lofgren P. Personalized PageRank Estimation and Search: A Bidirectional Approach / P. Lofgren, S. Banerjee, A. Goel // *Proceedings of the Ninth ACM International Conference on Web Search and Data Mining*. – 2016. – New York: Association for Computing Machinery. – P. 163-172. <https://doi.org/10.1145/2835776.2835823>.
5. Segmenting Two-Sided Markets / B. Banerjee, S. Gollapudi, R. Kollias, K. Munagala // *Proceedings of the 26th International Conference on World Wide Web*. – Republic and Canton of Geneva, CHE. – 2017. – P. 63-72. <https://doi.org/10.1145/3038912.3052578>.
6. Mauge K. Structuring e-commerce inventory / K. Mauge, K. Rohanimanesh, J.-D. Ruvini // *Proc. of the 50th Annual Meeting of the Association for Computational Linguistics*. – 2012. – P. 805-814. <https://www.aclweb.org/anthology/P12-1085>.
7. A Hybrid Framework for Product Normalization in Online Shopping / L. Wang, R. Zhang, C. Sha, X. He, A. Zhou // In: Meng W., Feng L., Bressan S., Winiwarter W., Song W. (eds) *Database Systems for Advanced Applications. DASFAA 2013, Lecture Notes in Computer Science*. – Vol. 7826. – Springer. – 2013. – P. 370-384. https://doi.org/10.1007/978-3-642-37450-0_28.
8. Towards Improving the Search Quality on the Trading Platforms / O. Cherednichenko, M. Vovk, O. Kanishcheva, M. Godlevskiy // In: S. Wrycza, J. Maslankowski (Eds): *11th SIGSAND/PLAIS 2018, LNBIP 333* – Springer. – 2018. – P. 21-30. https://doi.org/10.1007/978-3-030-00060-8_2.
9. Studying Items Similarity for Dependable Buying on Electronic Marketplaces / O. Cherednichenko, M. Vovk, O. Kanishcheva, M. Godlevskiy // *Proc. 2nd Int. Conf. On Computational Linguistics and Intelligent Systems (COLINS), Volume I: Main Conference (Lviv, Ukraine, June 25-27, 2018)*. – CEUR-WS. – 2018. – Vol. 2136. – P. 78-89. <http://ceur-ws.org/Vol-2136/10000078.pdf>.
10. *Statistics for business and economics* / D.R. Anderson, D.J. Sweeney, T.A. Williams et al. – South-Western: Cengage Learning, 2012. – 1090 p.

References

1. Turban, E., Whiteside, J., King, D. and Outland, J. (2017), *Introduction to Electronic Commerce and Social Commerce*, Springer, 413 p. <https://doi.org/10.1007/978-3-319-50091-1>.
2. Dinerstein, M., Einav, L., Levin, J. and Sundaresan, N. (2018), Consumer Price Search and Platform Design in Internet Commerce, *American Economic Review*, No. 108, pp. 1820-1859. <https://doi.org/10.1257/aer.20171218>.
3. Hortacsu, A., Babur, De los S., Matthijs, W. (2012), Testing Models of Consumer Search using Data on Web Browsing and Purchasing Behavior, *American Economic Review*, No. 102(6), pp. 2955-2980. <https://doi.org/10.1257/aer.102.6.2955>.
4. Lofgren, P., Banerjee, S. and Goel, A. (2016), Personalized PageRank Estimation and Search: A Bidirectional Approach, *Proceedings of the Ninth ACM International Conference on Web Search and Data Mining*, Association for Computing Machinery, New York, pp. 163-172. <https://doi.org/10.1145/2835776.2835823>.
5. Banerjee, B., Gollapudi, S., Kollias, R. and Munagala, K. (2017), Segmenting Two-Sided Markets, *Proceedings of the 26th International Conference on World Wide Web*, Republic and Canton of Geneva, CHE, pp. 63-72. <https://doi.org/10.1145/3038912.3052578>.
6. Mauge, K., Rohanimanes, K. and Ruvini, J.-D. (2012), Structuring e-commerce inventory, *Proc. of the 50th Annual Meeting of the Association for Computational Linguistics*, pp. 805-814. <https://www.aclweb.org/anthology/P12-1085>.
7. Wang, L., Zhang, R., Sha, C., He, X. and Zhou, A. (2013), A Hybrid Framework for Product Normalization in Online Shopping, *Database Systems for Advanced Applications. DASFAA 2013, Lecture Notes in Computer Science*, Vol. 7826, Springer, pp. 370-384. https://doi.org/10.1007/978-3-642-37450-0_28.
8. Cherednichenko, O., Vovk, M., Kanishcheva, O. and Godlevskiy, M. (2018), Towards Improving the Search Quality on the Trading Platforms, *11th SIGSAND/PLAIS 2018, LNBIP 333*, pp. 21-30. https://doi.org/10.1007/978-3-030-00060-8_2.
9. Cherednichenko, O., Vovk, M., Kanishcheva, O. and Godlevskiy, M. (2018), Studying Items Similarity for Dependable Buying on Electronic Marketplaces, *Proc. 2nd Int. Conf. On Computational Linguistics and Intelligent Systems (COLINS), CEUR-WS*, Vol. 2136, pp. 78-89. <http://ceur-ws.org/Vol-2136/10000078.pdf>.
10. Anderson, D.R., Sweeney, D.J. and Williams, T.A. (2012), *Statistics for business and economics*, Cengage Learning, South-Western, 1090 p.

Надійшла до редколегії 06.02.2020

Схвалена до друку 10.03.2020

Відомості про авторів:

Чередніченко Ольга Юріївна

кандидат технічних наук доцент
доцент кафедри Національного технічного
університету “Харківський політехнічний інститут”,
Харків, Україна
<https://orcid.org/0000-0002-9391-5220>

Information about the authors:

Olga Cherednichenko

Candidate of Technical Sciences Associate Professor
Senior Lecturer of National Technical
University “Kharkiv Polytechnic Institute”,
Kharkiv, Ukraine
<https://orcid.org/0000-0002-9391-5220>

Янголенко Ольга Василівна

кандидат технічних наук
доцент кафедри Національного технічного
університету “Харківський політехнічний інститут”,
Харків, Україна
<https://orcid.org/0000-0001-7755-1255>

Olha Yanholenko

Candidate of Technical Sciences
Senior Lecturer of National Technical
University “Kharkiv Polytechnic Institute”,
Kharkiv, Ukraine
<https://orcid.org/0000-0001-7755-1255>

Ивашенко Оксана Віталіївна

аспірант
Національного технічного університету
“Харківський політехнічний інститут”,
Харків, Україна
<https://orcid.org/0000-0003-3636-3914>

Oksana Ivashchenko

Doctoral Student
of National Technical University
“Kharkiv Polytechnic Institute”,
Kharkiv, Ukraine
<https://orcid.org/0000-0003-3636-3914>

Матвеев Александр Николаевич

аспірант
Національного технічного університету
“Харківський політехнічний інститут”,
Харків, Україна
<https://orcid.org/0000-0001-5907-3771>

Oleksandr Matveiev

Doctoral Student
of National Technical University
“Kharkiv Polytechnic Institute”,
Kharkiv, Ukraine
<https://orcid.org/0000-0001-5907-3771>

**МОДЕЛИ ФОРМИРОВАНИЯ РЕКОМЕНДАЦИЙ
В ИНТЕЛЛЕКТУАЛЬНЫХ СИСТЕМАХ ЭЛЕКТРОННОЙ КОММЕРЦИИ**

О.Ю. Чердниченко, О.В. Янголенко, О.В. Ивашенко, А.Н. Матвеев

Результаты работы поисковых и фильтрационных механизмов современных систем электронной коммерции не всегда удовлетворяют требования пользователей, что проявляется в неточных и неполных рекомендациях товаров по поисковому запросу. Совершенствование качества рекомендаций для покупателей онлайн торговых платформ является актуальной задачей. Данная работа приводит модели формирования рекомендаций на основе методов кластерного анализа, которые позволяют сгруппировать схожие товары и похожих клиентов по их характеристикам. Приведены результаты эксперимента по формированию рекомендаций касательно приобретения рюкзаков для покупателей онлайн магазина спортивного оборудования.

Ключевые слова: E-commerce система, рекомендационная интеллектуальная система, кластеризация, меры расстояния.

MODELS OF RECOMMENDATIONS FORMATION IN THE INTELLIGENT E-COMMERCE SYSTEMS

O. Cherednichenko, O. Yanholenko, O. Ivashchenko, O. Matveiev

With the increase of the amount of information available on the web, it becomes difficult for users to work with it. This applies, in particular, to e-commerce systems which are storing millions of products' offers. That is why systems that help users navigate in a large amount of information have begun to play a big role. A person cannot analyze a lot of information, because it is difficult and requires a lot of time and physical effort. But with recommendation systems that can filter out a lot of information and provide the user with the information they need and recommendations they like this problem can be solved. The results of the search and filtering mechanisms of modern e-commerce systems do not always satisfy the requirements of users, which is reflected in inaccurate and incomplete recommendations of products with specific search queries. Improving the quality of recommendations for buyers of online trading platforms is an urgent task.

The goal of this work is to improve the quality of e-commerce product search results by using an intelligent approach to clustering users according to their preferences and the products themselves according to their similarity by their own attributes. This paper provides a model of recommendation formulation based on cluster analysis methods that allow grouping of similar products and similar customers by their characteristics. The suggested algorithm uses Euclidian similarity measure between clusters' objects. The results of the experiment on forming recommendations for the purchase of backpacks for buyers of online sports equipment store are presented. The advantage of the developed models is that they use as an input a normalized set of products with unified representation of product attributes and their values. This allows to increase the precision and recall of search algorithms. The obtained results can be used in developing of new collaborative filtering algorithms and can increase the quality of recommendations and improve users' experience while interacting with the e-commerce system.

Keywords: E-commerce system, intelligent recommendation system, clustering, measures of distance.