

Технічні науки

УДК 004.2

**Шевченко Денис Олегович**

*студент*

*Харківського національного університету радіоелектроніки*

**Науковий керівник:**

**Лєсна Наталя Советівна**

*професор кафедри програмної інженерії*

*Харківський національний університет радіоелектроніки*

**ДОСЛІДЖЕННЯ МЕТОДІВ ПЕРСОНАЛІЗОВАНОГО  
ОБСЛУГОВУВАННЯ КОРИСТУВАЧІВ WEB-РЕСУРСІВ З МЕТОЮ  
ПІДБОРУ НАЙВИГІДНІШОЇ ПРОПОЗИЦІЇ**

***Анотація.** У статті автором досліджено основні принципи персоналізації контенту за умов ведення діяльності у сфері електронної комерції.*

*Стаття звертає увагу на методи застосування фундаментальної теорії ймовірностей до знаходження релевантності кожної окремої товарної позиції відповідно до користувача web-ресурсу.*

***Ключові слова:** персоналізація, пропозиція, електронна комерція.*

**Проблема і актуальність дослідження.** Актуальність дослідження методів персоналізованого обслуговування зумовлена необхідністю аналізу потреб та побажань кожного окремого користувача web-ресурсів щодо перегляду інформації, яка надає певну пропозицію, наприклад, товар, послугу або подію. Даній темі приділяється особлива увага у сфері електронної

комерції, адже якісна персоналізація дозволяє підвищити продажі товарів інтернет-магазинів на десятки, та навіть сотні відсотків. Тим не менш, на теперішній час навіть окремі лідери сфери електронної комерції не мають достатньо розвиненої системи персоналізації контенту для своїх користувачів, а існуючі системи для персоналізації зазвичай не включають до аналізу дані ресурсу, що за своєю природою швидко змінюються, наприклад, доступність товару на складі, або актуальність події в реальному часі. При цьому персоналізація із залученням існуючих систем має на меті виведення релевантного контенту для певного сегменту користувачів, але не аналізує прибутковість вихідної пропозиції. Вирішення питання прибутковості, тобто кореляція побажань користувача із ціною, доступністю та доречністю певного товару, залишається основною задачею менеджерів з продажу.

**Огляд поточного стану об'єкту дослідження.** Проблема персоналізації пропозицій бере початок із давніх принципів основ торгівельної діяльності. У своєму первинному вигляді персоналізація поставала питанням: "Яким чином комерсантові вплинути на окремого потенційного клієнта, щоб останній погодився на пропозицію?". Слід відзначити, що найбільший розвиток теоретичних аспектів персоналізації відбувся у другій половині ХХ ст. Цей прогрес було зумовлено невпинним розвитком психології, комерції, та маркетингу. В результаті множина успішних тактик персоналізованого обслуговування, що застосовувалася у межах окремих груп підприємців та компаній, почала активно досліджуватися та набувати формалізованої бази. Зокрема слід відзначити роботу Роберта Чалдіні "Психологія впливу", де досліджується поступливість особистості за умов існування аксіоми "Матеріального інтересу" та шести основних методів персоналізованого впливу на рішення: послідовність, взаємний обмін, соціальне доведення,

авторитет, прихильність, та дефіцит [1].

**Мета дослідження.** Поточна робота містить дослідження проблем персоналізації та аналізу прибутковості з метою їх оптимального вирішення у виді єдиної програмної моделі. Не зважаючи на те, що станом на поточний час персоналізація активно застосовується у сфері інформаційних технологій та має добре сформовану теоретичну базу, зазвичай її практичне застосування обмежено розподіленням користувачів на специфічні кластери та відповідним підбором контенту під кожний окремий сегмент відвідувачів. Новизною досліджуваної моделі є відповідність дій підбору персоналізованого контенту до приросту прибутку від пропозиції. Таким чином, застосування відповідної системи до основних програмних модулів інтернет-магазину має значно підвищити дохідність організації за рахунок постачання тих товарів, які найбільше цікавлять кожного окремого користувача, і водночас матимуть найліпший ефект з продажу для компанії.

**Теоретична база та питання.** Основою поточного дослідження є визначення взаємозв'язків та причинно-наслідкових залежностей між принципами персоналізації даних та виведенням найприбутковішої пропозиції для організації-провайдера товарів або послуг. Спочатку розглянемо предметну область загалом, поступово поглинаючи у проблемний домен і виділення питань до дослідження. Розгляд предметної області проводиться на базі загальних відомостей про структуру уявного інтернет-магазину.

Будь який ресурс електронної комерції має певну ієрархію товарних одиниць. Під коренем ієрархії зазвичай мається на увазі один, або декілька так званих каталогів. Декілька каталогів можуть представляти дочірні організації, або, наприклад, окремі варіанти товарних одиниць для певної

країни. Під каталогом знаходиться ієрархія категорій, яка відповідає за навігацію користувача до кінцевого уточнення розділу, де знаходяться продукти. Так як один і той же продукт може мати необмежену кількість варіацій за розміром, кольором, країною походження, тощо, в термінології електронної комерції вводиться поняття SKU - ідентифікатор товарної позиції, або одиниця обліку запасів. Саме визначення множини SKU є необхідним типом даних для проведення персоналізованого підбору контенту. На практиці до представлення SKU-даних додається певне навантаження атрибутами варіації, адже ця сутність є кінцевою точкою деталізації товару, і саме за допомогою введення цих атрибутів стає можливим виділення схожості та відмінності однієї товарної позиції від іншої. Крім атрибутів, що представляють властивості товарної позиції, кожна одиниця обліку має цінову характеристику. Зазвичай ціни зберігаються та керуються окремо від даних каталогу і мають декілька варіацій: базова ціна, ціна розпродажу, та вартість. Також ціна може корегуватися у залежності від впливу глобальних або місцевих знижок, які при зниженні цінової характеристики на окремі позиції підвищують загальний ефект від продажів. Окрім якісних та цінових показників товару, кожна одиниця обліку має відображення даних щодо наявності. В контексті електронної комерції ця інформація називається інвентарем, і має найшвидшу природу змін - при великому навантаженні більшість операцій із залученням змін до інвентаря намагаються зробити асинхронними. Товарна позиція може бути присутньою у обсязі достатньому для повноцінної пропозиції, обмеженої пропозиції (товар закінчується), та повністю відсутньою. Так як інтернет-магазин представляється клієнт- серверною системою, кожен користувач явно або неявно відповідає програмному профілю, який існує в контексті бізнес-

логіки системи. Варіація користувацьких профілів здійснюється за допомогою двох підходів: “спостерігання у якості співбесідника”, тобто безпосередній запит клієнтських даних (вік, стать, явні побажання), та “спостерігання у якості учасника”, тобто фіксація дій у системі без участі користувача: відстеження переходу до певного розділу, перегляду сторінки певний час, купівлі товарної позиції з певними характеристиками.

Наведені сутності, а саме вибірка якісних даних про товарну позицію, цінова характеристика, інвентар, та профіль користувача є необхідною та достатньою кількістю даних для проведення персоналізації вибірки контенту із попереднім аналізом релевантності та прибутковості пропозиції. Під релевантністю надалі розуміється відповідність позиції побажанням клієнта, тобто схожість даних до відображення із очікуваннями користувача.

**Основною гіпотезою**, що визначає подальше дослідження, є залежність приросту активності створення замовлень у межах інтернет-магазину та збільшення середньої вартості кожного окремого замовлення від ступеню релевантності відображуваної інформації.

Основні питання, що постають у досліджуваному домені:

- «які товарні позиції мають найвищий ступінь релевантності для певного користувача у певний момент часу?»;
- «які товарні позиції можливо реалізувати для певного користувача у певний момент часу враховуючи обмеженість бюджету?»;
- «які товарні позиції найвигідніше реалізувати з точки зору організації-постачальника враховуючи обмеженість запасів складу?»;
- «який взаємозв'язок можна встановити між релевантністю товарної позиції для користувача та вигодою реалізації для постачальника?»;
- «який спосіб сегментації даних користувачів та товарних позицій має

бути застосований за умов великого навантаження?»;

- «який критерій оптимальності необхідно застосовувати при верифікації результуючої вибірки даних?».

**Методи дослідження.** Беручи до уваги існуючі програмні реалізації персоналізованого обслуговування, слід відзначити застосування алгоритмів кластеризацій. Наприклад, при існуванні визначеної множини користувачів та товарних позицій із відповідними атрибутами вхідне програмне навантаження можна представити у вигляді матриці, яка відображає відповідність між користувачем та певною одиницею обліку. Прикладом алгоритму, що доцільно застосувати до подібної ситуації, є колаборативна фільтрація за допомогою ко-кластеризації. Втім, при існуванні десятків тисяч товарних одиниць та мільйонів користувачів поодинокі персоналізація із залученням кластерного аналізу призводить до суттєвої похибки та значних витрат на обчислення даних.

Альтернативним методом знаходження відповідей на попередньо поставлені питання є використання фундаментальної теорії ймовірностей. Розглянемо формулу Баєса

$$P(A|B) = P(B|A)P(A)/P(B) \quad (1)$$

у загальному контексті. Припустимо, що питання персоналізації постає у обчисленні вірогідності того, що має місце подія А, за умови, коли подія В гарантовано сталася. За формулою Баєса ця вірогідність пропорційна добутку правдоподібності  $P(B|A)$  на апіорну вірогідність  $P(A)$ . Знаменник  $P(B)$  відображає нормувальний коефіцієнт [2].

**Інтерпретація результатів.** Застосовуючи формулу Баєса до конкретних питань персоналізації маємо наступні положення: вірогідність А відображатиме релевантність товарної одиниці для певного користувача

системи, тобто представлятиме вірогідність її включення до замовлення. Априорна вірогідність  $P(A)$  може бути представлена як загальна популярність товарної позиції, наприклад, відштовхуючись від динаміки продажів, новизни, витрат на маркетинг, помножена на рівень інвентаря, який спільний для всієї системи. Вірогідність  $P(B|A)$  приймає вид спільних подій відношення детермінованих питань щодо певної товарної одиниці на обмеженій множині відповідей. Питання може поставати до наявності у певної одиниці товару визначеного атрибута, який притаманний профілю поточного користувача, наприклад: "Колір цього об'єкту червоний?" При цьому знаменник  $P(B)$  відображатиме коефіцієнт вигідності видачі обчисленої вірогідності.

Постає додаткове завдання підбору питань до системи, відповіді на які приведуть до найшвидшого обчислення цільових товарів для відображення. Для обчислення таких питань пропонується використання інформаційної ентропії. Таким чином, на кожній ітерації обчислення буде відома кількість інформації, якої бракує для видалення невизначеності. Формула для обчислення інформаційної ентропії:

$$H[P(A_i|B, < Q_j, C_1 \rangle)]P(< Q_j, C_1 \rangle) + \dots + H[P(A_i|B, < Q_j, C_N \rangle)]P(< Q_j, C_N \rangle) \quad (2),$$

де значення  $C_1, \dots, C_N$  множина релевантності відповіді до поставленого запитання - від повної відповідності до нуля. Описаний підхід є базовим механізмом роботи експертних систем діагностики, визначення об'єктів за умов неможливості формалізації, а також слугує апаратом відомої гри Akinator [3].

**Висновки.** В результаті дослідження маємо описану модель для персоналізованого обслуговування користувачів web-ресурсів із обчисленням найвигіднішої пропозиції. Верифікація отриманої моделі може бути здійснена

за допомогою підрахунку різниці значень AOV (Average Order Value) та ОРН (Order Per Hour) за умов постійного навантаження на web-ресурс до і після впровадження системи. Слід зауважити, що описана система потребуватиме адаптаційного періоду за умов відсутності значного тестового навантаження одразу після впровадження, так як точність розрахунків напряду залежить від попереднього досвіду обчислень, тобто чим більше задач персоналізації постає перед системою, тим більш релевантними будуть результати в подальшому.

### **Література**

1. R. Cialdini Influence: The Psychology of Persuasion [Text]/ Robert Cialdini - HarperCollins, 2001. - 279 с.
2. E. T. Jaynes Probability Theory: The Logic of Science [Text]/ E. T. Jaynes - Cambridge University Press, 2003. -753 с.
3. Акіатор та математика / Електронний ресурс / URL: <http://geektimes.ru/post/84364>, 2010.