

УДК 004.42

Аліпа М. Н., Куріленко А. Р., Волощук С. О.

ОЦІНКА ЕФЕКТИВНОСТІ РЕКОМЕНДАЦІЙНИХ СИСТЕМ IPTV, ЩО ВРАХУЮТЬ НЕЯВНІ УПОДОБАННЯ КОРИСТУВАЧІВ

У роботі розглянуто оцінки ефективності рекомендаційних систем відеоконтенту, описані адаптації методу визначення подібності інформаційних одиниць за допомогою аналізу даних про уподобання користувачів. Метод визначення подібності реалізує підхід Item-Item колаборативної фільтрації, який є одним з найбільш популярних підходів до побудови сучасних рекомендаційних систем. Вихідними даними для колаборативної фільтрації при побудові рекомендаційних систем IPTV є дані про користувача активності при перегляді одиниць відеоконтенту. Для формування об'єктивних критеріїв оцінки ефективності рекомендаційних систем відеоконтенту враховується, що дані збираються як явним (оцінки, опитування, рейтинги), так і неявним чином (протоколювання дій користувачів). Метод визначення подібності дозволяє вирішити проблему холодного старту, тобто видачі рекомендацій в період відсутності докладної інформації про користувача, але при наявності неявних даних про маршрути інших відвідувачів системи IPTV. Серед об'єктивних показників роботи рекомендаційної системи для оцінки обрані точність рекомендацій і їх повнота, а також узагальнююча їх оцінка. Для оцінки економічного ефекту експлуатації рекомендаційної системи задана класифікація клієнтів по інтенсивності використання рекомендацій. Задана цільова функція економічного ефекту експлуатації рекомендаційної системи. З'ясовано економічний ефект і ефективність використання рекомендаційної системи. Обговорено критерії прийняття рішення про доцільність використання рекомендаційної системи: експлуатаційна вартість і економічний ефект за розрахунковий період. Обговорено критерії прийняття рішення про доцільність оптимізації рекомендаційної системи, що використовується, або впровадження нової із завданням серед них пріоритетів відповідно до цілей стратегічного планування з урахуванням ризиків адаптації і підгонки.

Ключові слова: рекомендаційна система, фільм, хостінг, оцінка ефективності, колаборативна фільтрація.

Постановка проблеми. Розробка і оцінка ефективності рекомендаційних систем (РС), мають ряд серйозних проблем, які мають великий вплив на якість використовуваних моделей і їх адаптацію, незважаючи на всю їх гнучкість і велику кількість різних методів. Основні з цих проблем: відсутність інформації про користувача, висока розрідженість в даних, холодний старт, «білі ворони», масштабованість, підроблені оцінки [1]. Оптимізацію існуючих РС необхідно вести в напрямку вирішення цих проблем з огляду на одночасно можливі проблеми адаптації впровадженень, що реалізуються. Для цього необхідно виділення об'єктивних критеріїв оцінки ефективності такої оптимізації, що дозволяють говорити про її доцільність і потенціал в майбутньому.

Аналіз останніх досліджень та публікацій. Основне завдання рекомендаційної системи інтернет телебачення - формування контенту, що максимально відповідає очікуванням, в тому числі неявним, конкретного користувача. Для вирішення цього завдання

в більшості сучасних рекомендаційних систем використовується один з двох базових підходів: колаборативна фільтрація (collaborative filtering, CF) і тематична фільтрація (content-based filtering, CbF) [2, 3]. Метод контентної фільтрації фокусується на виявленні об'єктів зі схожими характеристиками по відношенню до тих об'єктів, які вже зацікавили користувача. При цьому враховується модель поведінки користувача і характеристики (контент) об'єктів, що зацікавили його. При виробленні рекомендацій виявляються об'єкти зі схожими характеристиками. Для ефективної роботи методу контентної фільтрації, як правило, необхідний докладний опис характеристик об'єктів, а також відомості про конкретного користувача (наприклад, відповіді на конкретні питання в анкеті). В основі методу колаборативної фільтрації лежить припущення про консервативність користувацьких переваг (користувачі, однаково оцінюють певні об'єкти, швидше за все аналогічним чином будуть оцінювати і нові об'єкти з подібними характеристиками) [3]. По суті, рекомендації базуються на автоматичній співпраці безлічі користувачів і на виділення (методом фільтрації) тих користувачів, які демонструють схожі переваги або шаблони поведінки. Таким чином, метод колаборативної фільтрації виробляє рекомендації, засновані на моделі попередньої поведінки користувача і з урахуванням поведінки користувачів зі схожими характеристиками. Найбільшого поширення в сфері електронної комерції отримали рекомендаційні системи, що використовують такі варіанти реалізації методу колаборативної фільтрації, а також їх гібриди: колаборативна фільтрація за допомогою аналізу переваг груп користувачів зі схожими інтересами (User-User CF); колаборативна фільтрація за допомогою аналізу взаємозв'язків між об'єктами (Item-Item CF). Основними проблемами, пов'язаними з реалізацією та практичним використанням алгоритмів колаборативної фільтрації, є розрідженість даних, проблема холодного старту і масштабованість. Додатково до перерахованих проблем можна відзначити проблему обмеження різноманітності пропозицій. Рекомендаційні системи, що використовують колаборативну фільтрацію, схильні пропонувати товари, що вже користуються популярністю, що створює проблеми для просування нових товарів і послуг [4, 5].

У методі User-User CF визначається схожість між користувачами і в якості рекомендацій користувачеві видається n товарів, що саме часто купуються, k найбільш схожими на нього покупцями. Для оцінки ступеня схожості користувачів в плані їх переваг можуть використовуватися різні функції подібності (метрики). Найбільш популярними серед них є: евклідова відстань, косинусна міра, відстань Хеммінга, коефіцієнт кореляції Пірсона, коефіцієнт Танімото і деякі інші [4, 6]. Визначення рекомендацій методом User-User CF передбачає побудову матриці активності користувачів, кожен рядок якої описує дії конкретного користувача стосовно до конкретного об'єкта (категорія, фільм, послуга) на сайті. Дії користувачів можуть позначатися самими різними способами. Наприклад, це може бути бінарна інформація про відвідування або не відвідування заданого ресурсу даним користувачем, частота (або число) користувань ресурсу r користувачем u , вартість або рейтинг, проставлений користувачем u для ресурсу r і т.д. Таким чином, кожен рядок матриці активності являє собою вектор оцінок, які відповідають різним категоріям одиниць відеоконтенту (тематичний профіль користувача). Профіль користувача характеризує ступінь його інтересу до кожної групи фільмів. Для кожної пари «користувач-об'єкт (фільм, послуга, дія)» в матриці активності обчислюється міра близькості з використанням обраної метрики [7]. Для пошуку рекомендацій конкретного користувача на підставі його поведінкового профілю використовуються три основні підходи: заснований на сусідстві

(memory based), заснований на моделі (model based) і гібридний підхід (hybrid). У сучасних комерційних рекомендаційних системах найбільшого поширення набули гібридний підхід і підхід, заснований на використанні моделей (алгоритми кластеризації, байєсовські мережі довіри, латентні семантичні моделі) [3, 7]. Для виявлення груп користувачів зі схожими характеристиками часто використовуються різні алгоритми кластеризації. Метод Item-Item CF історично з'явився як альтернатива методу User-User CF, покликана підвищити продуктивність рекомендаційних систем для тих магазинів, де число покупців істотно перевищує кількість найменувань товарів в каталозі [6]. Проблеми User-User CF: «холодний старт» істотно знижує якість роботи РС внаслідок відсутності даних про переваги нових (або мало активних) користувачів; проблема частого поновлення даних про користувача активності (в разі компанії, що пропонує послуги IPTV мова йде про десятки тисяч активних клієнтів) різко знижує продуктивність рекомендаційної системи в цілому.

Базові об'єктивні показники оцінки ефективності РС: Precision, Recall і узагальнюючий їх - F_1 . Precision (точність) - це частка об'єктів X : $X = Y \cap Z$, які цікаві користувачеві, серед всіх рекомендованих об'єктів - Y при всіх об'єктах, цікавих користувачеві Z :

$$\text{Precision} = |X| / |Y| \tag{1}$$

Recall (повнота) - це частка об'єктів, які були рекомендовані системою - X , серед всіх об'єктів, цікаві користувачеві [3], схематично наведені на рисунку 1.

$$\text{recall} = |X| / |Z| \tag{2}$$

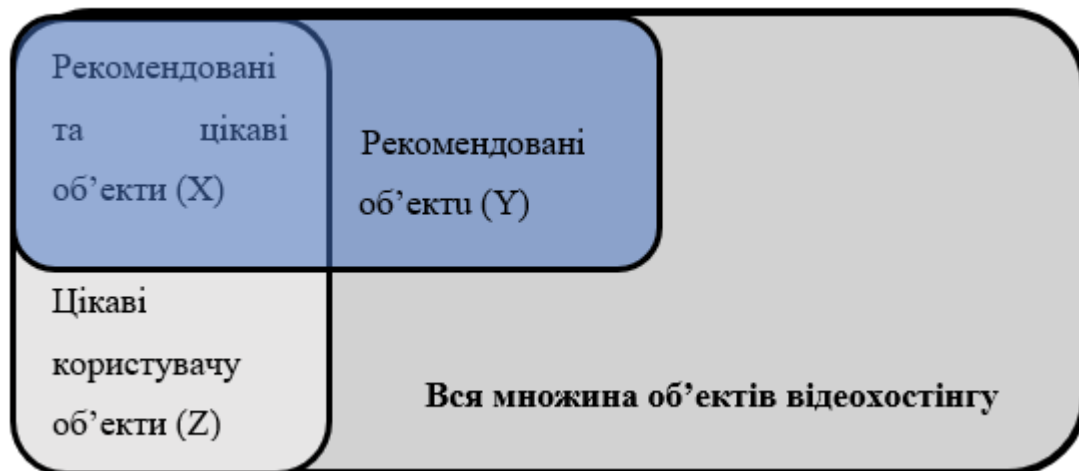


Рисунок 1 – Розрахунок метрик Recall та Precision

F_1 міра використовується для обліку відразу обох характеристик, так як можливі випадки, коли одна з метрик Precision чи Recall показує хорошу якість моделі, однак значення другої сильно просідає.

$$F_1 = 2 \frac{\text{precision} \cdot \text{recall}}{\text{precision} + \text{recall}} \tag{3}$$

Мета дослідження. Дослідження оцінки ефективності рекомендаційних систем, її адаптація і перевірка адекватності, розробка комбінованого підходу до побудови рекомендаційних систем, що забезпечує найбільш повне використання всіх даних про відеохостингу, пропонувані компанією пропонує послуги IPTV, і користувачів з метою вироблення рекомендацій, найбільш адекватно відображають їх очікування (пертіненість пропозиції). Науково практична новизна роботи полягає в ідеї комбінованого використання методів Item-Item CF і User-User CF, що дозволяє мінімізувати недоліки кожного з них та, ймовірно, домогтися більш високої якості роботи рекомендаційної системи в цілому. Такий підхід дозволяє виробляти прийнятні за якістю рекомендації в умовах, коли відомості про користувача перевагах відсутні, мінімальні або слабо інформативні. Для виявлення груп подібних товарів використовуються методи кластеризації, що дозволяє домогтися гарних показників якості та швидкодії в роботі алгоритму. З'ясовано економічний ефект і ефективність використання РС, та обговорено критерії прийняття рішення про доцільність використання РС відповідно до цілей стратегічного планування з урахуванням ризиків адаптації і підгонки РС.

Викладення основного матеріалу. Основна ідея методу Item-Item CF полягає в угрупованні інформаційних одиниць відеоконтенту, що мають подібні рейтинги користувачів. Рекомендації виробляються за наступним принципом: користувачеві оцінити об'єкт X високо буде запропонований об'єкт Y, який високо оцінили інші користувачі, також високо оцінили і об'єкт X. Використання методу Item-Item CF дозволяє підвищити якість рекомендацій для нових користувачів (немає критичної залежності від даних про уподобання користувача), а також значно підвищує продуктивність РС в разі, коли кількість користувачів істотно перевищує кількість об'єктів (характеристики об'єктів змінюються рідше). При цьому якість рекомендацій в середньому вище, ніж в разі використання підходу, заснованого на аналізі профілів. Для обчислення попарній близькості одиниць відеоконтенту можуть використовуватися ті ж метрики, що і у випадку з парами «користувач-об'єкт». Для пошуку рекомендацій на підставі матриці об'єктів часто використовуються вагові функції та методи регресійного аналізу [7].

Визначення подібності (кластерів) інформаційних одиниць по неявним уподобанням користувача реалізується при вирішенні завдання формування рекомендації з доречною інформацією в умовах недостатності знань про пристрасті користувачів, де пропонується використовувати метод, в основі якого лежить розрахунок близькості пар і подальше формування інформаційних кластерів на основі даних користувачів, послідовно переглядають декілька фільмів. При відсутності даних використовуються звичайні класифікатори з урахуванням ціни і параметрів об'єктів, список «Новинки», а також матрицю «З цим фільмом дивляться». При цьому підході явна участь користувачів у формуванні рейтингу фільмів не потрібна. Першим кроком алгоритму є побудова матриці подібності інформаційних одиниць, де і по вертикалі, і по горизонталі присутні всі одиниці відеохостингу. Заповнення матриці відбувається за таким правилом: якщо користувач послідовно переглянув два фільми, то вага подоби в матриці для цих двох товарів збільшується на 1. Матриць оброблена з метою виявлення груп інформаційних одиниць, які є близькими за своїми оцінками подібності методом кластеризації, випробуваному на тестовому масиві даних. Каталог відеохостингу містить більше 20 000 фільмів, які розбиті на 17 класів (12 кореневих класів, 5 підкласів). Щодня IPTV відвідують в середньому близько 10

000 відвідувачів і проводять на ньому близько 50 хвилин кожен. Вихідні дані охоплюють період в один квартал (IV квартал 2018 роки), в якому каталог фільмів був незмінний.

На їх основі побудована матриця подібності по всьому тимчасовому періоду. Матриця сильно розріджена, так як для багатьох пар фільмів оцінка подібності відсутня (тобто протягом аналізованого періоду часу користувачі не цікавилися деякими фільмами з каталогу). В результаті обробки матриці подібності побудована гістограма відстаней. Виділено кластери, серед яких є великі. Якість роботи алгоритму можна оцінити по референсній інформаційній одиниці, де в кластер подібності потрапили фільми одного жанру або побічно асоціюються з деякими акторами. Аналогічні результати показує і дослідження інших отриманих кластерів. Таким чином, метод визначення подібності інформаційних одиниць виконує покладені на нього завдання: формується рекомендація з фільмів, потенційно цікавих для невідомого відвідувача в даний конкретний момент часу.

Для розрахунку економічного ефекту розрахунковим періодом обраний 1 рік, для якого задається вартість C амортизації (після придбання) або оренди РС. Клієнти класифікуються за інтенсивністю використання РС, коли цей параметр монетизує потенціал задоволеності РС, як частку від LTV (Customer Lifetime Value, довічна цінність клієнта). Для простої поінформованості про наявність РС частка α (зі значенням 0.1 % причому в інтерфейсі відеохостингу передбачена можливість відключення стрічки з рекомендаціями за бажанням клієнта), в разі використання РС параметр β (на інтервалі $-2\% \div 5\%$) обчислюється відповідно з алгоритмом на стандартних оцінках задоволеності користувачів РС наведених в табл.1, добре корелюють з рейтингами, що виставляються клієнтами.

Таблиця 1 – Стандартні оцінки задоволеності клієнтів

	значення <i>Precision, Recall, F₁</i>			
	$\geq 0,75$	$\geq 0,5$	$\geq 0,25$	$\leq 0,25$
Оцінка	відмінно	добре	задовільно	погано

Для визначення успішності рекомендації задається коефіцієнт огляду за розрахунковий період d з наступною диференціацією: перегляд більше 50 % фільму - 1, більше 25 % - 0.5; більше 10 % - 0.25; менше 10 % - 0. Далі при множині обраних фільмів Z (де переглянутим вважається контент з $d \geq 0.5$) і рекомендованих - Y за розрахунковий період обчислюються оцінки точності і повноти представлених рекомендацій. $\forall z_i \in Z$ перевіряється чи $z_i \in Y$, для якого рекомендація вважається успішною і формується множина X , z_i при $d = 0.25$ в розрахунку β не беруть участь, z_i при $d = 0$ вважаються поганими рекомендаціями і по ним нараховуються пенальті у вигляді штрафної функції, що зміщує оцінку за шкалою β в сторону незадоволеності. Згідно виразів (1) та (2) обчислюються повнота та точність рекомендацій - $Pr(\text{precision})$ та $Rc(\text{recall})$, відповідно, та за виразом (3) розраховується метрика F_1 яка узагальнює показники Pr і Rc і шкалюється по області значень β . Перевірка адекватності об'єктивних оцінок може виконуватися після виключення «білих ворон» з оціночних множин у вигляді порівняння з суб'єктивними оцінками, отриманими у вигляді рейтингів фільмів проставляють користувачами. Після чого обчислюються цільові функції економічного ефекту РС - F_α та F_β по класах інтенсивності використання клієнтами РС у вигляді:

Інформаційні технології

$$F_{\alpha} = \sum LTV_i \cdot \alpha \tag{4}$$

$$F_{\beta} = \sum LTV_i \cdot \beta \tag{5}$$

Загальна цільова функція економічного ефекту від використання РС – F_0 , розраховується як сума цільових функцій F_{α} та F_{β} по класах інтенсивності використання клієнтами РС згідно виразів (5) та (6).

У таблиці 2 наведено фрагмент даних про 10 клієнтів з БД клієнтів із базовими характеристиками ID, LTV, абонплатою, даними про рекомендації та просмотр фільмів (Z, Y, X; Xd – кількість досмотрених фільмів), доповнені розрахунковими даними: penalty (в %), штрафи для РС за недогляд рекомендованого фільму – $p = (1-Xd) / X$; Re, Rc, F_1 – метрики (оцінки) якості роботи РС, розраховані згідно виразів (1)÷(3), природа оцінок представлена на рис.1; F_{1d} – скоригована метрика F_1 якості роботи РС з з урахуванням даних про догляд фільмів, F_0 – цільова функція монетизації «задоволеності» клієнта від користування РС.

Таблиця 2 – Наведені статистичні дані по користувачам РС, клієнтам провайдера

F ₀ , usd	ID	абон- плата, uah	LTV, usd	Z	Y	Xd	X	d	P, %	Re	Rc	F ₁	F _{1d}
0,71	10	180,00	88,16	144	96	48	55	7	12,73	0,38	0,57	0,46	0,40
-0,54	1	70,00	34,29	293	202	15	109	94	86,24	0,37	0,54	0,44	0,06
0,52	2	180,00	88,16	155	111	49	57	7	12,92	0,36	0,51	0,43	0,37
0,51	3	180,00	88,16	120	87	38	47	8	17,85	0,39	0,53	0,45	0,37
0,39	4	130,00	63,67	160	107	50	63	13	20,00	0,39	0,58	0,47	0,37
0,25	5	130,00	63,67	272	193	80	100	20	20,00	0,37	0,52	0,43	0,34
0,19	6	70,00	34,29	282	196	87	105	18	17,14	0,37	0,54	0,44	0,36
0,29	7	130,00	63,67	284	203	85	107	21	20,00	0,37	0,52	0,44	0,35
0,53	8	180,00	88,16	177	127	57	65	8	12,25	0,36	0,51	0,42	0,37
0,54	9	180,00	88,16	370	265	118	142	23	16,47	0,38	0,53	0,45	0,37

На рис. 2 наведено залежність економічних характеристик використання РС від чисельності клієнтів відеохостингу, N. F_0 – функція монетизації «задоволеності» клієнтів для однієї з найдешевших РС, пропонованих на ринку. Річна абонплата за використання РС буде близько 200 USD з можливістю більш частого оновлення БД фільмів. Точка беззбитковості досягається при кількості клієнтів, що користуються послугами відеохостингу, $N \geq 1300$. F_0' – функція монетизації «задоволеності» клієнтів для РС, яка буде обходитися компанії-провайдеру 1000 USD за рік. В цьому випадку беззбитковість досягається при кількості клієнтів $N \geq 2900$. На малюнку приведені порівняльні річні експлуатаційні вартості РС в 200, 1000 і 3000 USD у вигляді ліній C_1 , C_2 і C_3 , відповідно. Крива F_{notD} показує функцію монетизації «задоволеності» клієнтів РС з річною експлуатаційною вартістю C_2 без урахування догляду відеоконтенту клієнтами. На якісному рівні зрозуміло, що підвищення частоти недогляду відеоконтенту клієнтом збільшує рівень його «незадоволеності» РС в цілому, що повинно відобразитися в збільшенні негативного ефекту монетизації цільової функції. На рис. 2 видно, що кількісно економічний ефект відрізняється разуче. Модельні

відповідності обрані у вигляді адекватних емпіричних залежностей для наявних натурних даних і можуть вимагати доопрацювання за критерієм адекватності для інших вибірок клієнтської БД або при екстраполяційному прогнозуванні економічних показників в разі збільшення кількості клієнтів.

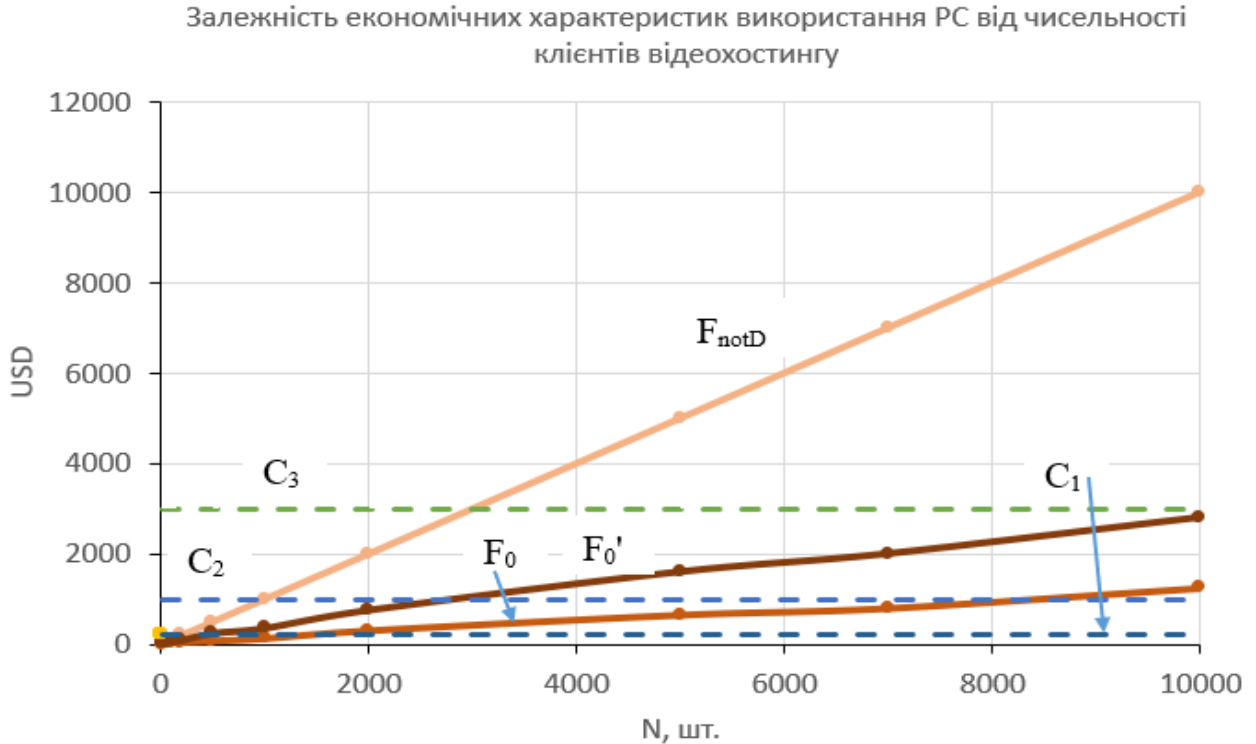


Рисунок 2 – Залежність економічних характеристик використання РС від чисельності клієнтів відеохостингу N:

F_0, F_0' – функція монетизації «задоволеності» клієнтів РС із річною експлуатаційною вартістю C_1, C_2 , відповідно; F_{notD} функція монетизації «задоволеності» клієнтів РС з річною експлуатаційною вартістю C_2 без урахування догляду відеоконтенту клієнтами. Приведено річні експлуатаційні вартості РС в 200, 1000 і 3000 USD у вигляді ліній C_1, C_2 і C_3 , відповідно

На рис. 3 наведено залежність показників економічної ефективності експлуатації РС - F_0, Δ_0, Φ_0 (показник фондovіддачі - $\Phi_0 = F_0 / C_0$) від чисельності клієнтів відеохостингу N при річній експлуатаційній вартості РС - C_1 . Чистий економічний ефект і фондovіддача (Δ_0, Φ_0 відповідно) виходять в «+» зі значень чисельності клієнтів від 1300 осіб. Рішення про доцільність використання РС приймається після порівняння за початковий (0-й) розрахунковий період експлуатаційної вартості РС – C_0 і цільової функції економічного ефекту РС – F_0 , де успіх буде при $\Delta_0 > 0$ при $\Delta_0 = F_0 - C_0$ з економічною ефективністю Φ_0 . Рішення про доцільність оптимізації РС, що використовується, або впровадженні нової РС приймається після порівняння проектних показників F', Δ', Φ' з відповідними існуючими F_0, Δ_0, Φ_0 із завданням серед них пріоритетів відповідно до критеріїв стратегічного планування, причому збільшення показників виражених в абсолютних величинах повинно бути мінімум в

3÷4 рази з урахуванням ризиків адаптації і підгонки РС. На рис.3 ця величина помічена червоним хрестиком. Чистий економічний ефект і фондovіддача F_0' , Δ_0' , Φ_0' при експлуатації РС при річній експлуатаційній вартості C_2 виходять в «+» зі значень чисельності клієнтів від 2900 осіб, а досягають відповідності критеріям оптимізації РС аналогічно наведеним вище для рис.3 при об'ємі клієнтури відеохостингу 30÷40 тис. осіб.

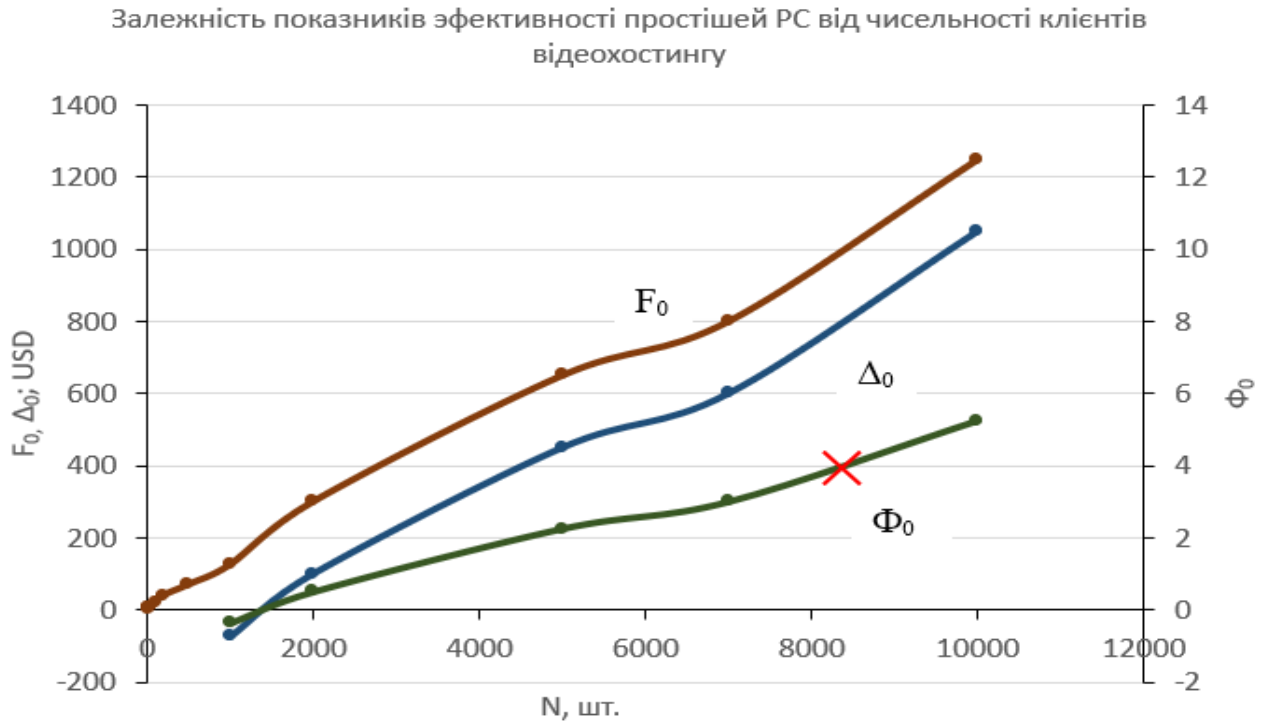


Рисунок 3 – Залежність показників ефективності рекомендаційної системи з річною експлуатаційною вартістю C_1 від чисельності клієнтів відеохостингу N :

F_0 - функція монетизації «задоволеності» клієнтів РС (тотожна кривій F_0 на малюнку 2), Δ_0 - економічний ефект від використання РС, Φ_0 - показник фондovіддачі ($\Phi_0 = F_0 / C_0$), відповідно. Червоним хрестиком помічена точка, де економічний ефект від використання РС Δ_0 мінімум в 3÷4 рази перевищує вартість експлуатації

ВИСНОВКИ

Персоналізація контенту на сьогодні є одним з напрямків, що активно розвиваються, для розміщення відеофайлів в ІТ-індустрії. Найважливішими результатами її застосування є збільшення залучення відвідувачів, поліпшення користувацького досвіду і підвищення конверсії. Персоналізація контенту реалізується за допомогою рекомендаційних систем. Сучасні рекомендаційні системи забезпечують обробку величезних масивів даних про активність користувача з метою формування прогнозів для конкретних користувачів в момент запиту. У роботі викладено метод визначення подібності одиниць відеоконтенту по неявним уподобанням користувача в рекомендаційних системах, який можна застосовувати для нових (або малоактивних) користувачів. При цьому в міру накопичення даних про переваги користувачів рекомендуються надавати більшу перевагу методам класу User-User

CF, які дають тим точніші прогнози чим більш докладні дані про користувача активності. Виконано оцінку ефективності рекомендаційних систем, її адаптація і перевірка адекватності. З'ясовано кількість клієнтів компанії, що представляє послуги відеохостінгу для досягнення відповідності критеріїв оптимізації РС.

Перелік використаних джерел:

1. Probabilistic memory-based collaborative filtering / K. Yu [et al.] // IEEE Transactions on Knowledge and Data Engineering. – 2004. – Vol. 16, N 1. – P. 56–69.
2. Подходы к повышению пертинентности информационного предложения в медиасервисах на основе обработки больших объемов данных / С. А. Филиппов, В. Н. Захаров, С. А. Ступников, Д. Ю. Ковалев // Аналитика и управление данными в областях с интенсивным использованием данных : XVII международная конференция DAMDID/RCDL'2015 (Обнинск, 13–16 октября 2015 г.). – Обнинск, 2015. – С. 224–228.
3. Aggarwal, C. C. Recommender Systems / C. C. Aggarwal. – [S. l.] : Springer, 2016. – 493 p.
4. Xiaoyuan, Su A Survey of Collaborative Filtering Techniques[Electronic resource] / Xiaoyuan Su, Taghi M. Khoshgoftaar // Advances in Artificial Intelligence. – 2009. – Vol. 2009. – Article ID 421425. – P. 1–19. – Mode of access: <https://downloads.hindawi.com/archive/2009/421425.pdf>
5. Fleder, D. Blockbuster Culture's Next Rise or Fall: The Impact of Recommender Systems on Sales Diversity / D. Fleder, K. Hosanagar // Management Science. – 2009. – Vol. 55, N 5, May. – P. 697–712.
6. Брейкин, Е. А. Рекомендательная система на основе коллаборативной фильтрации / Е. А. Брейкин // Молодой ученый. – 2015. – № 13. – С. 31–33.
7. Методы и модели анализа данных: OLAP и Data Mining / А. А. Барсегян, М. С. Курприянов, В. В. Степаненко, И. И. Холод. – СПб. : БХВ-Петербург, 2004. – 336 с.

Алипа М. Н., Куриленко А. Р., Волощук С. А.

ОЦЕНКА ЭФФЕКТИВНОСТИ РЕКОМЕНДАТЕЛЬНЫХ СИСТЕМ IPTV, УЧИТЫВАЮЩИХ НЕЯВНЫЕ ПОЛЬЗОВАТЕЛЬСКИЕ ПРЕДПОЧТЕНИЯ

В процессе работы рассмотрены оценки эффективности рекомендательных систем видеоконтента, описанные адаптации метода определения сходства информационных единиц с помощью анализа данных о предпочтениях пользователей. Метод определения сходства реализует подход Item-Item коллаборативной фильтрации, который является одним из самых популярных подходов к построению современных рекомендательных систем. Исходными данными для коллаборативной фильтрации при построении рекомендательных систем IPTV данные о пользовательской активности при просмотре единиц видеоконтента. Для формирования объективных критериев оценки эффективности рекомендательных систем видеоконтента учитывается, что данные собираются как явным (оценки, опросы, рейтинги), так и неявным образом (протоколирование действий пользователей). Метод определения сходства позволяет решить проблему холодного старта, то есть выдачи рекомендаций в период отсутствия подробной информации о пользователе, но при наличии

явных данных о маршрутах других посетителей системы IPTV. Среди объективных показателей работы рекомендательной системы для оценки выбраны точность рекомендаций и их полнота, а также обобщающая их оценка. Для оценки экономического эффекта эксплуатации рекомендательной системы задана классификация клиентов по интенсивности использования рекомендаций. Заданная целевая функция экономического эффекта эксплуатации рекомендательной системы. Выяснен экономический эффект и эффективность использования рекомендательной системы. Обсуждены критерии принятия решения о целесообразности использования рекомендательной системы: эксплуатационная стоимость и экономический эффект за расчетный период. Обсуждены критерии принятия решения о целесообразности оптимизации рекомендательной системы, используемой или внедрение новой с задачей среди них приоритетов в соответствии с целями стратегического планирования с учетом рисков адаптации и подгонки.

Ключевые слова: рекомендательная система, фильм, хостинг, оценка эффективности, коллаборативная фильтрация.

Alipa M. N., Kurilenko A. R., Voloschuk S. A.

ESTIMATING THE EFFICIENCY OF IPTV RECOMMENDATION SYSTEMS TAKING INTO ACCOUNT THE IMPLICIT CUSTOM PREFERENCES

The paper reviews the effectiveness evaluations of recommender video content systems, describes the adaptations of the method for determining the similarity of information units by analyzing user preference data. The similarity determination method implements the Item-Item approach of collaborative filtering, which is one of the most popular approaches to building modern recommendation systems. The initial data for collaborative filtering when constructing IPTV recommendation systems are data on user activity when viewing units of video content (information clusters). To form objective criteria for evaluating the effectiveness of recommender systems of video content, it is taken into account that data is collected both explicitly (ratings, surveys, ratings) and implicitly (logging user actions). The method of determining similarity allows to solve the problem of cold start, i.e. issuing recommendations in the absence of detailed information about the user, but if there are implicit data on the routes of other visitors to the system. Among the objective indicators of the recommendation system for evaluation, the accuracy of the recommendations and their completeness, as well as a generalizing assessment, were selected. To assess the economic effect of the operation of the recommendation system, a classification of customers according to the intensity of the use of recommendations is given. The objective function of the economic effect of the exploitation recommendatory is given. The economic effect and the effectiveness of using the recommendation system are clarified. Criteria for deciding on the advisability of using a recommender system are discussed: operational cost and economic effect for the settlement period. The criteria for deciding on the advisability of optimizing the recommender system used or introducing a new one with setting priorities among them in accordance with the goals of strategic planning taking into account the risks of adaptation and adjustment are discussed.

Keywords: recommendation system, film, hosting, performance evaluation, collaborative filtering.

Рекомендовано до публікації: доц., к.т.н., Сімкін А.І.

Стаття надійшла 28.08.2020 р.