

Міністерство освіти і науки України
Одеський національний технологічний університет
Кафедра комп'ютерної інженерії



**ПОЯСНЮВАЛЬНА ЗАПИСКА
ДО КВАЛІФІКАЦІЙНОЇ РОБОТИ**

на тему Дослідження і розробка рекомендаційних
(назва кваліфікаційної роботи згідно наказу ОНТУ)
систем на основі асоціативних правил

Здобувача Здор В.О.
(прізвище, ініціали)

2 курсу 561 а групи

Керівники: к.т.н., доц. Сахарова С.В.
(посада, прізвище та ініціали)
ст. викл. Журнова Т.М.
(посада, прізвище та ініціали)

Консультанти: к.т.н., доц. Нєнов О.
(посада, прізвище та ініціали)
д.е.н., проф. Басюркіна Н
посада, прізвище та ініціали)

Кваліфікаційна робота допускається до захисту

Рішення кафедри від 30. 11 2023 р., протокол № 3

Завідувач кафедри комп. інженерії Сергій АРТЕМЕНКО
(назва кафедри) (підпис) (Ім'я ПРИЗВИЩЕ)

Одеса – 2023 рік

ОДЕСЬКИЙ НАЦІОНАЛЬНИЙ ТЕХНОЛОГІЧНИЙ УНІВЕРСИТЕТ

Факультет комп'ютерної інженерії, програмування та кіберзахисту
Кафедра комп'ютерної інженерії
Ступінь вищої освіти магістр
Спеціальність 123 «Комп'ютерна інженерія»
Освітня програма Комп'ютерні системи та мережі

ЗАТВЕРДЖУЮ

Зав. кафедри комп'ютерної інженерії
Сергій АРТЕМЕНКО
« 30 » 11 2023 року

ЗАВДАННЯ НА КВАЛІФІКАЦІЙНУ РОБОТУ ЗДОБУВАЧА

Здора Владислава Олександровича

1. Тема роботи Дослідження і розробка рекомендаційних систем на основі асоціативних правил

Затверджена наказом університету від « 30 » 11 2022 р., наказ № 884-03

2 Термін здачі здобувачем закінченої роботи 25 листопада 2023 р.

3. Вихідні дані роботи

1. Методи аналізу даних

2. Методи факторного аналізу

3. Асоціативні правила.

4. Перелік питань, які потрібно розробити

1. Вступ. 2. Аналіз принципів розробки рекомендаційних систем.

3. Аналіз підходів заснованих на перевагах користувачів.

4. Реалізація методу, розробка прикладу системи відбору відеоматеріалів на основі асоціативних правил

5. Економічні розрахунки. 6. Охорона праці. 7. Загальні висновки.

5. Перелік графічного матеріалу (з точним зазначенням обов'язкових креслень)

Інтерактивна презентація (мета, об'єкт, предмет дослідження, постановка задачі;

Методи, що використовуються при розробці рекомендаційних систем; Опис вихідних даних
Асоціативні правила; Приклад колаборативної фільтрації; Використовувані метрики;

Приклад відбору відеоматеріалів; Висновки.)

6. Консультанти по роботі, із зазначенням розділів роботи, що стосуються їх

Розділ	Консультант	Підпис, дата	
		Завдання видав	Завдання прийняв
Економіка	Барсюкіна Н. Й., д.е.н., проф.		
Охорона праці	Нєнов О. Л., к.т.н., ст. викл.		
Нормоконтроль	Сахарова С.В., к.т.н., доц.		

7. Дата видачі завдання 10.10.2023

Керівники

Світлана САХАРОВА

Тетяна Журнова

Завдання прийняв до виконання

Владислав ЗДОР

КАЛЕНДАРНИЙ ПЛАН

№	Назва етапів кваліфікаційної роботи	Термін виконання етапів роботи	Примітка
1.	Дослідження об'єкту.	05.09.2023	
2.	Дослідження технології.	05.10.2023	
3.	Постановка завдання. Визначення вихідних даних.	07.11.2023	
4.	Аналіз підходів заснованих на перевагах користувачів	07.11.2023	
5.	Реалізація методу, розробка прикладу системи відбору відеоматеріалів на основі асоціативних правил	20.11.2023	
6.	Підготовка техніко-економічної частини	25.11.2023	
7.	Підготовка розділу охорони праці	25.11.2023	
8.	Оформлення пояснювальної записки	25.11.2023	
9.	Оформлення графічної частини та лістингу	25.11.2023	
10.			

Здобувач-дипломник _____ Владислав ЗДОР

Керівник роботи _____ Світлана САХАРОВА

Несу відповідальність за ідентичність електронного та друкованого варіантів кваліфікаційної роботи, даю згоду на обробку персональних даних та не заперечую проти розміщення кваліфікаційної роботи на офіційних web-ресурсах ОНТУ.

Підтверджую, що в кваліфікаційній роботі відсутні порушення норм академічної доброчесності.

Здобувач-дипломник Владислав ЗДОР _____

АНОТАЦІЯ

У рамках роботи розглянуто популярний метод інтелектуального аналізу даних, що полягає у побудові асоціативних правил, що ґрунтуються на частих множинах. Також аналізуються методи засновані на колаборативній фільтрації та запропоновано алгоритм, особливості якого є орієнтація на рекомендації користувачам, з унікальними перевагами.

Порівняння методів проводитиметься з використанням даних MovLens.

Представлено опис використовуваних даних. Виконана формалізація використовуваних підходів, а саме, асоціативні правила, колаборативна фільтрація та описано використаний алгоритм. Порівняння результатів роботи вищеописаних алгоритмів і висновки.

Розроблено та проаналізовано приклад роботи рекомендаційної системи відбору відеоматеріалів на основі асоціативних правил та колаборативної фільтрації, що забезпечило підвищення швидкодії та якості надання рекомендації.

Також виконано економічний розрахунок та розглянуто питання охорони праці.

Ключові слова: рекомендаційні системи, асоціативні правила.

ABSTRACT

As part of the work, a popular method of intelligent data analysis, which consists in building associative rules based on frequent sets, is considered. Methods based on collaborative filtering are also analyzed, and an algorithm is proposed, the features of which are orientation to user recommendations, with unique advantages.

Comparison of the methods will be done using Movielens data.

A description of the data used is presented. The used approaches are formalized, namely, associative rules, collaborative filtering, and the used algorithm is described. Comparison of the results of the above algorithms and conclusions.

An example of the recommendation system for selecting video materials based on associative rules and collaborative filtering was developed and analyzed, which ensured an increase in the speed and quality of providing recommendations.

An economic calculation was also performed and the issue of labor protection was considered.

Keywords: recommendation systems, associative rules.

.

ЗМІСТ

ВСТУП.....	8
РОЗДІЛ 1 АНАЛІЗ ПРИНЦИПІВ РОЗРОБКИ РЕКОМЕНДАЦІЙНИХ СИСТЕМ	11
1.1 Аналіз стану питання в галузі розробки рекомендаційних систем та застосування методів аналізу даних	11
1.2 Основні визначення	13
РОЗДІЛ 2 АНАЛІЗ ПІДХОДІВ ЗАСНОВАНИХ НА ПЕРЕВАГАХ КОРИСТУВАЧІВ	23
2.1 Опис вихідних даних.	23
2.2 Опис підходу, що є прикладом колаборативної фільтрації.....	25
2.3 Часті множини та асоціативні правила	27
РОЗДІЛ 3 РЕАЛІЗАЦІЯ МЕТОДУ, РОЗРОБКА ПРИКЛАДУ СИСТЕМИ ВІДБОРУ ВІДЕОМАТЕРІАЛІВ НА ОСНОВІ АСОЦІАТИВНИХ ПРАВИЛ ...	36
3.1 Оцінка якості та тестування рекомендаційної системи	36
3.2 Розробка прикладу рекомендаційної системи відбору відеоматеріалів на основі асоціативних правил	40
РОЗДІЛ 4 ЕКОНОМІЧНІ РОЗРАХУНКИ ПРОЕКТУ.....	52
РОЗДІЛ 5 ОХОРОНА ПРАЦІ.....	69
ВИСНОВКИ.....	84
СПИСОК ВИКОРИСТАНИХ ДЖЕРЕЛ.....	86
ДОДАТКИ.....	87

Владислав ЗДОР

					КРМ.КІ.1.884-03.2.3							
Змн.	Арк.	№ докум.	Підпис	Дата								
Розробив					Дослідження і розробка рекомендаційних систем на основі асоціативних правил				Літ.	Арк.	Акрушів	
Перевірів		Світлана САХАРОВА										
Рецензент												
Нормоконтроль												
Затвердив		Сергій АРТЕМЕНКО										
гр. 561, ОНТУ												

ВСТУП

В даний час існує проблема надлишку інформації, яка спричиняє труднощі у виділенні релевантних знань, топіків, об'єктів з доступної множини. Релевантність в інформаційній науці та інформаційному пошуку означає ступінь відповідності знайденого документа або набору документів інформаційним потребам користувача. Через це ускладнюється процес вибору користувачами тих товарів та послуг в Інтернеті, які підходять їм більше, ніж всі інші доступні.

Ця проблема має особливе місце в галузі електронної комерції, коли користувачі потребують рекомендації щодо того, що їм може сподобатися при виборі товару чи послуги. Переваги від цього мають як користувачі, коли знаходять саме те, що їм потрібно, так і Інтернет-магазини, сайти, які пропонують свої послуги та товари.

Вже давно існують способи більш точно прогнозувати інтерес користувача до різних товарів. Щоб зробити рекомендацію користувачеві необхідна історія його інтересів до цього моменту. Проте, можна зіткнутися з "проблемою холодного старту" якщо історія користувача відносно невелика. Це може бути з багатьох причин, наприклад, це новий користувач, іноді буває, що стандартні методи недостатньо добре прогнозують переваги такого користувача.

Більше того, зустрічаються такі користувачі, переваги яких рідко зустрічаються в інших. З такими складно працювати, тому що вони губляться на тлі загальної маси, але вони представляють інтерес, оскільки таких користувачів може бути все-таки багато.

У рамках роботи розглянуто популярний метод інтелектуального аналізу даних, що полягає у побудові асоціативних правил, що ґрунтуються на частих множинах. Також аналізуються методи засновані на колаборативній фільтрації та запропоновано алгоритм, особливості якого є орієнтація на рекомендації користувачам, з унікальними перевагами.

					KPM.KI.0.884-03.2.3	Арк.
						8
Змн.	Арк.	№ докум.	Підпис	Дат		

Рекомендаційні системи нині відіграють значну роль багатьох сферах, що з обробкою великих обсягів інформації, наприклад онлайн-магазини і кінотеатри, сервіси для прослуховування музики. Як правило, побудова рекомендацій ґрунтується на аналізі інформації про уподобання користувача, отримані в минулому. Часто така інформація подається у вигляді матриці користувачів-об'єктів. Однак у деяких ситуаціях інформація про користувача може бути недоступна, наприклад, під час першого або анонімного відвідування сервісу. Подібна постановка завдання характерна для окремого класу рекомендаційних систем – рекомендаційних систем на основі сесій (Session-based Recommender Systems, SBRS). На відміну від класичних підходів до побудови рекомендаційних систем, SBRS спираються тільки на дані про поточну сесію користувача, що дозволяє отримувати інформацію про його переваги, що швидко змінюються. Метою рекомендацій, як правило, є передбачення наступного об'єкта, на який зверне увагу користувач, або безліч таких об'єктів у поточній сесії. У цій статті наводиться огляд основних алгоритмів, які використовуються при побудові SBRS. Також наводиться огляд та порівняння фреймворків, призначених для розробки таких рекомендаційних систем.

Загальна ідея полягає у створенні рекомендаційної системи, яка на основі ретроспективного аналізу, тобто попередніх виборів різних користувачів, відповідає, які об'єкти треба рекомендувати новому користувачу, таким чином, щоб досягти максимального перетину чи співпадіння з тим, що його дійсно зацікавить.

Як правило, для рекомендаційних систем використовуються алгоритми навчання з учителем. Тобто є навчальна вибірка, що представляє собою безліч об'єктів, безліч користувачів і безліч рейтингів чи оцінок проставлених кожним з користувачів кожному з об'єктів. Під словом рейтинги може матися на увазі як просто факт вибору об'єкта користувачем, так і ступінь його задоволеності цим об'єктом.

Порівняння методів проводитиметься з використанням даних Movielens. Незважаючи на те, що дані вже багато років використовуються не тільки

					KPM.KI.0.884-03.2.3	Арк.
						9
Змн.	Арк.	№ докум.	Підпис	Дат		

дослідниками, які займаються рекомендаційними системами, а й тими, хто вивчає вже напрацьовані результати, вони однаково залишаються актуальними.

MovieLens - це веб-система рекомендацій та віртуальна спільнота, яка рекомендує користувачам дивитися фільми на основі їх переваг, використовуючи спільну фільтрацію оцінок фільмів та оглядів фільмів учасників. Він містить близько 11 мільйонів оцінок приблизно до 8500 фільмів. MovieLens був створений у 1997 р. GroupLens Research, дослідницькою лабораторією факультету комп'ютерних наук та інженерії Університету Міннесоти, з метою збору дослідницьких даних за персоналізованими рекомендаціями.

Розробка та аналіз рекомендаційних систем є досить актуальною задачею, оскільки кожен користувач, як споживач, стикається з проблемою вибору товарів чи послуг в Інтернеті і кожен постачальник товарів чи послуг є заінтересованим в збільшенні обсягу продажів та в формуванні бази постійних клієнтів. Задоволення потреб клієнтів збільшує прибутки постачальників.

Об'єктом дослідження є рекомендаційні системи.

Предметом дослідження є методи аналізу даних такі як асоціативні правила, колаборативна фільтрація.

Мета роботи – дослідження рекомендаційних систем, що використовують асоціативні правила.

Наукова новизна – дослідження та застосування методів факторного аналізу при розробці рекомендаційних систем.

Постановка **задачі** полягає у наступному: представлено опис використовуваних даних. Виконана формалізація використовуваних підходів, а саме, асоціативні правила, колаборативна фільтрація та описано використаний алгоритм. Порівняння результатів роботи вищеописаних алгоритмів і висновки.

					KPM.KI.0.884-03.2.3	Арк.
						10
Змн.	Арк.	№ докум.	Підпис	Дат		

РОЗДІЛ 1

АНАЛІЗ ПРИНЦИПІВ РОЗРОБКИ РЕКОМЕНДАЦІЙНИХ СИСТЕМ

1.1 Аналіз стану питання в галузі розробки рекомендаційних систем та застосування методів аналізу даних

Проблема знаходження релевантної інформації актуальна протягом останніх кількох десятиліть. За цей час було створено безліч алгоритмів штучного інтелекту, покликаних вирішити її. Існує кілька основних способів побудови рекомендаційних систем.

Одним з найбільш ранніх та успішних є метод колаборативної фільтрації, Описаний у статті [1]. Ідея його полягає в тому, щоб для нового користувача знайти "сусідів", щоб його переваги збігалися. Потім, продукти, що сподобалися "сусідам" будуть рекомендовані цьому користувачеві, оскільки вони найімовірніше сподобаються йому. Автори оцінюють як впливають різні способи формування "сусідів", що залежать від розміру сусідства, розмірності даних на результати та порівнюють їх із продуктивністю асоціативних правил. Незважаючи на те, що колаборативна фільтрація дає хороші результати на практиці, автори перераховують найістотніші недоліки цього підходу.

Одна з проблем – це масштабованість результатів. Найчастіше рекомендаційні системи мають працювати на даних, у яких можуть бути мільйони як користувачів, так і об'єктів. У такому разі процес формування "сусідів" може тривати багато часу. При цьому в багатьох системах потрібен миттєвий результат. Таким чином, виникає потреба у великих обчислювальних потужностях.

Наступна проблема полягає в якості рекомендацій, пропонованих користувачам. У рекомендаційних системах, як і в багатьох інших алгоритмах, є два типи помилок: помилково позитивні, це рекомендації, які не сподобалися користувачеві, і помилково негативні, це об'єкти, які могли б сподобатися користувачеві, але не були порекомендовані. У таких областях, як електронна

					KPM.KI.0.884-03.2.3	Арк.
						11
Змн.	Арк.	№ докум.	Підпис	Дат		

комерція, важливо заслужити довіру користувача. Тобто якщо кількість помилково позитивних рекомендацій перевищуватиме якісь емпірично встановлені межі, то користувач надалі навряд чи прислухатиметься до рекомендацій.

Найчастіше ці дві проблеми конкурують між собою, тому що чим менше часу йде на пошук сусідів, тим більш масштабований алгоритм і тим гірше його точність. Тому в колаборативної фільтрації важливо враховувати ці обмеження одночасно, щоб результати вийшли практичними та доцільними.

Один із найпопулярніших підходів інтелектуального аналізу даних, вже зазначених вище, полягає у знаходженні асоціативних правил, що ґрунтуються на частих множинах. Такі алгоритми формування частих множин, як Apriori, Eclact, FPgrowth докладно описані в [2], [3], а також там вказані етапи створення рекомендацій. Перевага даного джерела полягає в тому, що автори описують не тільки алгоритми, а й те, якими можуть бути дані, плюси та мінуси підходів, а також усі наводять детальні приклади, на яких можна простежити всі особливості алгоритмів знаходження. частих множин. Також асоціативними правилами займалися [4]. У своїй статті автори запропонували алгоритм, для якого не потрібне визначення значення мінімальної підтримки для формування правил.

Тобто цей параметр налаштовується автоматично. Також автори стверджують що ео дозволяє не шукати всі можливі асоціативні правила, що задовольняють мінімальній підтримці, що може значно зменшити час роботи алгоритму.

Не можна не відзначити ще один популярний та успішний спосіб скласти РС, а саме матричну факторизацію в цілому та SVD. У 2006 році було змагання Netflix Prize, яке спровокувало великий прогрес у галузі колаборативної фільтрації та залучило багатьох дослідників. Єгуда Корен та Роберт Белл виграли цей конкурс і запропонували спосіб, який показав вражаючі результати, після успішної участі автори продовжили розробляти свою методику та описав її в [5]. Авторами описується підхід колаборативної

					KPM.KI.0.884-03.2.3	Арк.
						12
Змн.	Арк.	№ докум.	Підпис	Дат		

фільтрації, який аналізує відносини між користувачами та взаємозалежностями серед продуктів, виявлення нових взаємозв'язків між користувачами та об'єктами, не вимагаючи створення явних профілів.

Колаборативна фільтрація застосовується у двох основних напрямках. Перший це метод "сусідів" у якому обчислює відносини між об'єктами або між користувачами. Другий підхід є прикладом моделі неявних факторів, який намагається інтерпретувати рейтинги, характеризуючи об'єкти та користувачів, тоді нові фактори можуть бути виведені з якихось закономірностей у рейтингах. Матрична факторизація є найвдалішою реалізацією моделі прихованих факторів. Для свого дослідження автори використовують сингулярне розкладання матриць (SVD) як базовий алгоритм. У статті автори наголошують на успішності підходу в умовах, коли багато користувачів поставили малу кількість рейтингів і запропонували включати додаткові джерела інформації про користувачів у дослідження.

1.2 Основні визначення

Рекомендаційні системи — це комплекс сервісів та програм, що аналізує уподобання користувачів та намагається передбачити, що може їх зацікавити.

Рекомендаційні системи мають широку сферу застосування та призначення, тому їх можна поділяти на групи за різними параметрами: предметом, метою, джерелом рекомендації, користувальницьким контекстом і ступенем персоналізації. Наприклад, мета систем, які рекомендують фільми, книги, новини та інший контент – залучити користувача до споживання нової інформації на конкретній платформі. В цьому випадку пропозиція контенту базується не тільки на контексті дій самого користувача, але й на новизні інформації, оцінках експертів та популярності в інших користувачів. Інакше працюють системи, які радять придбати товари повсякденного попиту. Їхнє завдання — не тільки запропонувати нові продукти, виходячи з контексту дій користувачів, а й нагадати про товари, що часто купуються, за наявності історії покупок. Основна мета - підвищити середній чек та лояльність покупця.

					KPM.KI.0.884-03.2.3	Арк.
						13
Змн.	Арк.	№ докум.	Підпис	Дат		

Від поставлених цілей та необхідні характеристики залежить вибір методів реалізації.

Рекомендаційні системи — програми (веб-сервіс), які намагаються передбачити, які об'єкти (фільми, музика, книги, новини, веб-сайти) будуть цікавими для користувача, маючи певну інформацію про його профіль.

Дві основні стратегії створення рекомендаційних систем - фільтрація на основі змісту та колаборативна фільтрація.

При фільтрації на основі змісту створюються профілі користувачів та об'єктів, профілі користувачів можуть включати демографічну інформацію або відповіді на певний набір питань, профілі об'єктів можуть містити назви жанрів, імена акторів, імена виконавців та іншу атрибутивну інформацію залежно від типу об'єкта. Наприклад, у Music Genome Project музичний аналітик оцінює кожен композицію за сотнями різних музичних характеристик, які можуть використовуватися для виявлення музичних уподобань користувача.

При колаборативній фільтрації використовується інформація про поведінку користувачів у минулому - наприклад, інформація про покупки або оцінки. У цьому випадку немає значення, з якими типами об'єктів ведеться робота, але при цьому можуть враховуватися неявні характеристики, які складно було б врахувати при створенні профілю. Основна проблема цього типу рекомендаційних систем — «холодний старт»: відсутність даних про користувачів, що недавно з'явилися в системі, або об'єктах.

У процесі роботи рекомендаційні системи збирають дані про користувачів, використовуючи поєднання явних та неявних методів. Приклади явного збору даних:

- запит користувача оцінки об'єкта за диференційованою шкалою;
- запит у користувача ранжування групи об'єктів від найкращого до найгіршого;
- пред'явлення користувачеві двох об'єктів із питанням про те, який із них кращий;
- пропозиція створити список об'єктів, улюблених користувачем.

					KPM.KI.0.884-03.2.3	Арк.
						14
Змн.	Арк.	№ докум.	Підпис	Дат		

Приклади неявного збору даних:

- спостереження за тим, що оглядає користувач в інтернет-магазинах чи базах даних іншого типу;
- ведення записів про поведінку користувача онлайн;
- відстеження вмісту комп'ютера користувача.

Рекомендаційні системи порівнюють однотипні дані від різних людей та обчислюють список рекомендацій для конкретного користувача. Деякі приклади їх комерційного та некомерційного використання наведені у статті про колаборативну фільтрацію. Для обчислення рекомендацій використовується граф інтересів.

Рекомендаційні системи — зручна альтернатива пошуковим алгоритмам, оскільки дозволяють виявити об'єкти, які можуть бути знайдені останніми. Цікаво, що рекомендаційні системи часто використовують пошукові машини для індексації незвичайних даних.

Навчання асоціативним правилам або пошук асоціативних правил - це метод машинного навчання на базі правил виявлення зв'язків, що цікавлять користувача, між змінними у великій базі даних. Метод пропонується встановлення сильних правил, виявлених у базі даних з допомогою деяких заходів цікавості. Цей заснований на правилах підхід також генерує нові правила в міру аналізу додаткових даних. Кінцевою метою, виходячи з досить великого набору даних, допомогти машині імітувати виділення людських ознак і створити можливість знаходження абстрактних асоціацій з нових некласифікованих даних.

Крім аналізу ринкового кошика, асоціативні правила використовуються у багатьох областях, включаючи Web mining, виявлення вторгнень, безперервне виробництво та біоінформатику. На відміну від виявлення послідовнісних шаблонів, навчання асоціативним правилам зазвичай не враховує порядок елементів усередині транзакції або по транзакціях.

Асоціативні правила дозволяють знаходити закономірності між пов'язаними подіями. Прикладом такої закономірності є правило, що вказує, що

					KPM.KI.0.884-03.2.3	Арк.
						15
Змн.	Арк.	№ докум.	Підпис	Дат		

з події X слідує подія Y з певною ймовірністю. Встановлення таких залежностей дає можливість знаходити дуже прості та інтуїтивно зрозумілі правила.

Вперше завдання пошуку асоціативних правил було запропоновано для знаходження типових шаблонів покупок, що здійснюються в супермаркетах, тому іноді його ще називають аналізом ринкового кошика (market basket analysis).

Визначення. Нехай:

$I = \{i_1, i_2, i_3, \dots, i_n\}$ - множина (набір) товарів, які називаються елементами. Нехай:

D - множина транзакцій, де кожна транзакція

T - це набір елементів з $I \subseteq T \subseteq I$.

Кожна транзакція є бінарним вектором, де $t[k]=1$, якщо i_k елемент присутній у транзакції, інакше $t[k]=0$. Ми говоримо, що транзакція T містить X , деякий набір елементів із I , якщо $X \subset T$. Асоціативним правилом називається імплікація: $X \Rightarrow Y$ де $X \subset I, Y \subset I$ та $X \cap Y = \emptyset$. Правило $X \Rightarrow Y$ має підтримку s (support), якщо $s\%$ транзакцій з D , містять $X \cup Y$, $supp(X \Rightarrow Y) = supp(X \cup Y)$.

Іншими словами, метою аналізу є встановлення таких залежностей: якщо в транзакції зустрівся певний набір елементів X , то на підставі цього можна зробити висновок про те, що інший набір елементів Y також має з'явитися в цій транзакції. Встановлення таких залежностей дає нам можливість знаходити дуже прості та інтуїтивно зрозумілі правила.

Алгоритми пошуку асоціативних правил призначені для знаходження всіх правил XY , причому підтримка і достовірність цих правил повинні бути вищими за деякі наперед певні пороги, які називаються відповідно мінімальною підтримкою (minsupport) і мінімальною достовірністю (minconfidence).

Завдання знаходження асоціативних правил розбивається на дві підзадачі:

Знаходження всіх наборів елементів, які задовольняють поріг minsupport. Такі набори елементів називаються такими, що часто зустрічаються.

					KPM.KI.0.884-03.2.3	Арк.
						16
Змн.	Арк.	№ докум.	Підпис	Дат		

Генерація правил наборів елементів, знайдених згідно з п.1. з достовірністю, що задовольняє поріг minconfidence.

Було запропоновано багато алгоритмів для створення асоціативних правил.

Декілька алгоритмів добре відомі, це Apriori, Eclat і FP-Growth, але вони роблять лише половину роботи, оскільки вони призначені для відшукування найпоширеніших наборів об'єктів. Потрібно зробити ще один крок після того, як найпоширеніші набори знайдені в базі даних.

Один із перших алгоритмів, які ефективно вирішують подібний клас завдань, — це алгоритм APriori . Крім нього, були розроблені й інші алгоритми: DHP, Partition, DIC та інші.

Значення параметрів мінімальна підтримка та мінімальна достовірність вибираються таким чином, щоб обмежити кількість знайдених правил. Якщо підтримка має велике значення, то алгоритми будуть знаходити правила, добре відомі аналітикам або настільки очевидні, що немає сенсу проводити такий аналіз.

З іншого боку, низьке значення підтримки веде до генерації величезної кількості правил, що, звісно, потребує суттєвих обчислювальних ресурсів. Проте більшість цікавих правил перебуває саме за низького значення порога підтримки. Хоча занадто низьке значення підтримки веде до створення статистично необґрунтованих правил.

Пошук асоціативних правил зовсім не очевидне завдання, як може здатися на перший погляд. Одна з проблем - алгоритмічна складність при знаходженні наборів елементів, що часто зустрічаються, т.к. зі зростанням числа елементів у $I(|I|)$ експоненційно зростає кількість потенційних наборів елементів.

Приклади застосування

Аналіз ринкового кошика (market basket analysis). Алгоритм виявляє типові шаблони покупок і товарів, що спільно купуються. Отримані результати дозволяють оптимізувати асортимент та його розміщення у торгових залах, покращити управління запасами, збільшити обсяги продажу за рахунок пропозиції клієнтам супутніх товарів.

					KPM.KI.0.884-03.2.3	Арк.
						17
Змн.	Арк.	№ докум.	Підпис	Дат		

Крос-продаж (cross-sell) та продаж з підвищенням ціни (up-selling). Алгоритм дозволяє на підставі шаблонів споживчої поведінки виявити клієнтів, схильних до відгуків на персональні пропозиції та крос-продажів. Це дає можливість формувати пропозиції товарів та послуг ефективніше; забезпечується індивідуальний підхід обслуговування клієнтів.

Директ мейл (direct mail) – пряме адресне розсилання рекламних пропозицій потенційним та існуючим покупцям – є високоефективним, простим та дешевим маркетинговим інструментом. Для збільшення кількості відгуків на листи необхідно робити ретельний відбір об'єктів для розсилки послань, чому сприяє розглянутий алгоритм.

При пошуку асоціативних правил завдання було суттєво спрощено. По суті, все зводилося до того, присутній у транзакції елемент чи ні. Тобто. якщо розглядати випадок ринкового кошика, ми розглядали два стану: куплений товар чи ні, проігнорувавши, наприклад, інформацію у тому, скільки було куплено, хто купив, характеристики покупця тощо. І можна сказати, що розглядали «булівські» асоціативні правила. Якщо взяти будь-яку базу даних, то кожна транзакція складається з різних типів даних: числових, категоріальних і т.д. Для обробки таких записів та вилучення чисельних асоціативних правил було запропоновано алгоритм пошуку.

Приклад чисельного асоціативного правила: [Вік: 30-35] та [Сімейний стан: одружений] [Місячний дохід: 1000-1500 умовних одиниць].

Крім описаних вище асоціативних правил є непрямі асоціативні правила, асоціативні правила з запереченням, тимчасові асоціативні правила для подій пов'язаних у часі та інші.

Як було сказано, завдання пошуку асоціативних правил вперше було представлено для аналізу ринкового кошика. Асоціативні правила ефективно використовуються в сегментації покупців з поведінки при покупках, аналізі переваг клієнтів, плануванні розташування товарів у супермаркетах, крос-продажах, адресному розсиланні. Проте сфера застосування цих алгоритмів не обмежується лише однією торгівлею. Їх також успішно застосовують і в інших

					KPM.KI.0.884-03.2.3	Арк.
						18
Змн.	Арк.	№ докум.	Підпис	Дат		

галузях: медицині, для аналізу відвідувань веб-сторінок (Web Mining), для аналізу тексту (Text Mining), для аналізу даних з перепису населення, в аналізі та прогнозуванні збоїв телекомунікаційного обладнання тощо.

Алгоритм Apriori

Алгоритм Apriori використовує стратегію пошуку завширшки для підрахунку об'єктів та використовує функцію генерації кандидата, яка заснована на властивості низхідного замикання підтримки.

Алгоритм Eclat

Алгоритм Eclat (або ECLAT від Equivalence Class Transformation = Перетворення Класів Еквівалентності) є алгоритмом пошуку в глибину, заснованому на перетині множин. Алгоритм придатний як для послідовного, так і для паралельного виконання з властивостями локального поліпшення.

Алгоритм FP-зростання

Алгоритм FP призначений для виявлення шаблонів, що часто зустрічаються.

При першому проході алгоритм підраховує зустрічність об'єктів (пари атрибут-значення) у наборах і запам'ятовує їх у «таблиці заголовків». При другому проході алгоритм будує структуру FP-дерева шляхом вставки екземплярів. Об'єкти в кожному екземплярі повинні бути впорядковані за зменшенням їх частоти в наборі, так що дерево може бути оброблене швидко. Об'єкти в кожному екземплярі, які не досягають мінімального порогу, відкидаються. Якщо багато екземплярів мають загальними більшість об'єктів, що часто зустрічаються, FP-дерево забезпечує високе стиснення близько до кореня дерева.

Рекурсивна обробка цієї версії стиснення зростання великих об'єктів головного набору призначається прямо, а чи не генерується кандидати з наступною перевіркою на повну базу. Зростання починається знизу таблиці заголовків шляхом знаходження всіх екземплярів, що відповідають цим умовам. Створюється нове дерево з лічильниками, одержуваних з вихідного дерева та відповідними набору екземплярів, які залежать від атрибуту, і кожен

					KPM.KI.0.884-03.2.3	Арк.
						19
Змн.	Арк.	№ докум.	Підпис	Дат		

вузол отримує суму лічильників його нащадків. Рекурсивне зростання припиняється, коли не залишилося об'єктів, які задовольняють мінімальний поріг підтримки, і робота продовжується на інших елементах заголовків вихідного FP-дерева.

Коли рекурсивний процес завершено, все більші набори об'єктів з мінімальним покриттям знайдено і починається створення асоціативного правила.

AprioriDP

AprioriDP використовує динамічне програмування в аналізі найпоширеніших наборів об'єктів. Принцип роботи - виняток генерації кандидата як у FP-дереві, але алгоритм запам'ятовує лічильники підтримки не в дереві, а в специфічній структурі.

Заснований на контексті алгоритм пошуку асоціативних правил

CBPNARM — це алгоритм, розроблений у 2013 році, для виявлення асоційованих правил на базі контексту. Алгоритм використовує контекстну змінну, з урахуванням якої значення підтримки набору об'єкта змінюється і основі цього правила переносяться у безліч правил.

Алгоритми на основі безлічі вузлів

FIN, PrePost і PPV - це три алгоритми, засновані на множинах вузлів. Вони використовують вузли в кодуванні FP-дерева для представлення наборів об'єктів і підтримують стратегію пошуку в глибину для виявлення наборів об'єктів, що часто зустрічаються, за допомогою «перетину» наборів вузлів.

Процедура ASSOC методу GUHA

GUHA це загальний метод аналізу даних, який має теоретичні основи.

Процедура ASSOC є методом GUHA, який шукає загальні асоціативні правила за допомогою швидких операцій над бітовими рядками. Асоціативні правила, виявлені цим методом, більш загальні, ніж отримані методом Apriori, наприклад, «об'єкти» можуть бути пов'язані як кон'юнкцією, так і диз'юнкцією і зв'язок між лівою частиною та правою частиною правила не обмежена

					KPM.KI.0.884-03.2.3	Арк.
						20
Змн.	Арк.	№ докум.	Підпис	Дат		

встановленням мінімальних значень підтримки та довіри як у методі Apriori - може бути використана довільна комбінація заходів цікавості.

Пошук OPUS

OPUS є ефективним алгоритмом виявлення правил, який, на відміну багатьох альтернатив, вимагає обмеження ні монотонності, ні антимонотонності, як-от у мінімумі підтримки. Пошук OPUS є базовою технологією в популярній системі пошуку асоціацій Magnum Opus.

Мультиасоціативні правила (Multi-Relation Association Rules, MRAR), це асоціативні правила, в яких кожен об'єкт може мати кілька зв'язків. Ці зв'язки показують непрямі зв'язки між сутностями. Розглянемо наступне мультиасоціативне правило, в якому перший член складається з трьох відносин живе в, поруч і вологий: «Двоє, хто живе в місці, яке знаходиться поряд з містом з вологим кліматом, і молодше 20 років => їхній стан здоров'я хороший». Такі асоціативні правила можна отримати з даних RDBMS або семантичних даних Інтернету.

Засноване на контексті асоціативні правила є видом асоціативних правил. Стверджується, що ці правила точніші в аналізі асоціативних правил і працюють шляхом розгляду прихованої змінної, названої контекстною змінною, яка змінює кінцевий набір асоціативних правил, залежно від значень контекстних змінних. Наприклад, орієнтація на купівельну корзину в аналізі ринкового кошика відображає дивні результати на початку місяця. Це може бути спричинене контекстом, таким як видача зарплати на початку місяця.

Контрастне навчання (Contrast set learning) є видом асоціативного навчання. Контрастне навчання використовує правила, які суттєво відрізняються у їхньому розподілі за підмножинами.

Виважене навчання класам (англ. Weighted class learning) є іншим видом асоціативного навчання, в якому ваги можуть бути призначені класам, щоб перевести фокус на конкретні спірні питання для результатів дослідження даних.

					KPM.KI.0.884-03.2.3	Арк.
						21
Змн.	Арк.	№ докум.	Підпис	Дат		

Виявлення шаблонів високого порядку (англ. High-order pattern discovery) сприяє вилученню шаблонів високого порядку чи асоціативних подій, властивих даним складного реального світу.

Виявлення K-оптимальних шаблонів забезпечує альтернативу до стандартного підходу до навчання асоціативних правил, де потрібно, щоб кожен шаблон часто з'являвся в даних.

Наближений аналіз наборів, що часто зустрічаються (англ. Approximate Frequent Itemset mining) є ослабленою версією аналізу часто зустрічаються наборів, в якій допускається, щоб деякі з об'єктів в деяких рядках дорівнювали 0.

Узагальнені асоціативні правила (Generalized Association Rules) - ієрархічна класифікація

Кількісні асоціативні правила (Quantitative Association Rules) — категоріальні та кількісні дані.

Інтервальні асоціативні правила (англ. Interval Data Association Rules) містять дані, розбиті на інтервали, наприклад, вік з інтервалом в 5 років.

Виявлення послідовнісних шаблонів (Sequence pattern mining) виявляє підпослідовності, які є загальними для більш ніж minsup послідовностей в базі даних, де значення minsup встановлюється користувачем. Послідовність – це впорядкований список транзакцій.

Кластеризація підпростору (англ. Subspace Clustering), специфічний вид кластеризації даних високої розмірності, у багатьох випадках також ґрунтується на властивості низхідного замикання для специфічних моделей кластерів.

Wapmr постачається як частина комплекту ACE для аналізу даних. Система дозволяє навчання асоціативним правилам для реляційних правил першого ладу.

					KPM.KI.0.884-03.2.3	Арк.
						22
Змн.	Арк.	№ докум.	Підпис	Дат		

РОЗДІЛ 2

АНАЛІЗ ПІДХОДІВ ЗАСНОВАНИХ НА ПЕРЕВАГАХ КОРИСТУВАЧІВ

2.1 Опис вихідних даних

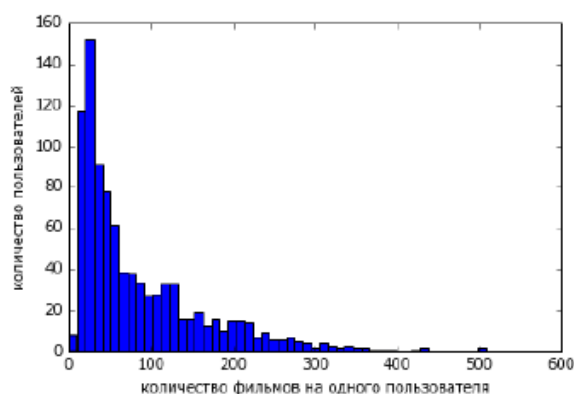
Перед використанням алгоритмів гарною практикою є процес розуміння особливостей даних, що покращує уявлення про очікувані результати. Для аналізу були використані дані MovieLens, що містять рейтинги 1500 фільмів, поставлені приблизно 1000 користувачів ними. При цьому кожен користувач оцінив щонайменше 20 фільмів. Вся вибірка буде розбита на навчальну та тестову щодо 80% та 20% відповідно. Оскільки надалі для застосування алгоритмів будуть потрібні бінарні дані, тобто таблиця з нулів та одиниць, необхідно якимось чином привести рейтинги до цього подання. Один із простих способів це зробити - це застосувати граничне шкалювання. Тобто призначити поріг, і замість значень не менше за нього ставити 1, а нижче 0. У такому разі одиницю можна інтерпретувати як факт задоволеності фільмом, а зворотне нуль.

Якщо ж поріг знизити до нуля, то 1 означатиме, що фільм буде переглянуто, і жодного висновку про задоволеність не буде зроблено. Це цілком реальна ситуація, оскільки задоволеність може залежати від такого великого спектру факторів, наприклад, симпатії користувача до жанру, акторів, режисера тощо. У практичних дослідженнях значення порога буде використано як параметр, що варіюється.

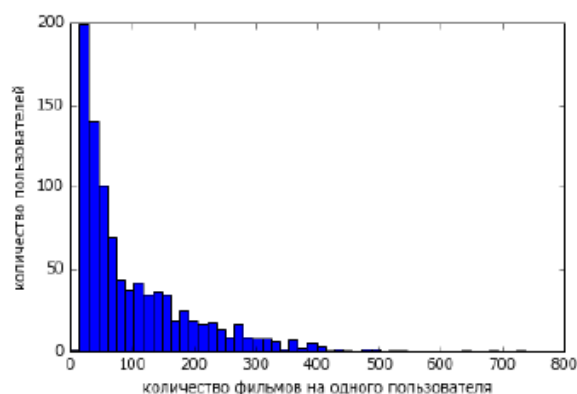
Для того, щоб краще розуміти особливості аналізованих даних, оцінимо як багато фільмів позитивно оцінено кожним з користувачів. Варто відразу зазначити, що якщо йдеться про позитивну оцінку, це означає, що обраний якийсь поріг не рівний 0, і позитивна оцінка - значення вище за поріг, негативна оцінка - нижче за поріг. На рисунку 1 а) представлена гістограма кількості фільмів одного користувача, тобто скільки фільмів було оцінено хоч якось кожним з користувачів. Якщо поріг відмінний від 0, і наприклад, дорівнює 3, то

					KPM.KI.0.884-03.2.3	Арк.
						23
Змн.	Арк.	№ докум.	Підпис	Дат		

гістограма з таким значенням параметра наведена на рисунку 1 б). З графіків видно, що в обох випадках в основному користувач оцінив 20-30 фільмів, і приблизно стільки ж позитивно. Однак це не дає повного уявлення про те, як саме розподілені дані та як між собою перетинаються самі фільми. Для цього оцінимо скільки користувачів припадає на кожен із фільмів, залежно від порога. На рисунку 2 а) поріг дорівнює 3, на рисунку б) поріг дорівнює 0. На обох графіках видно, що більшість фільмів вибрано дуже малою кількістю користувачів. Ця особливість наводить на думку, що методи повинні якимось знаходити спільність між якимись фільмами навіть якщо немає жодного користувача, який би переглянув їх усі. Більше того, ця особливість даних надалі вплине на продуктивність описуваних підходів.

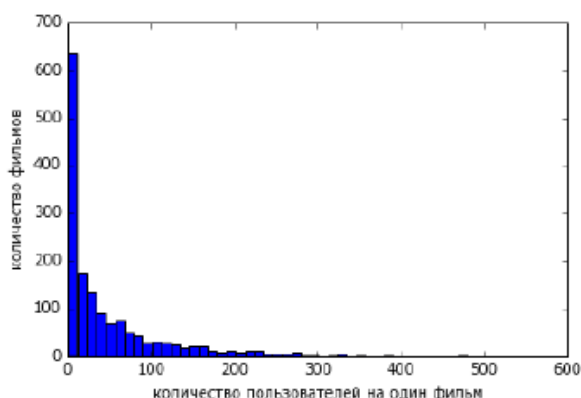


а)

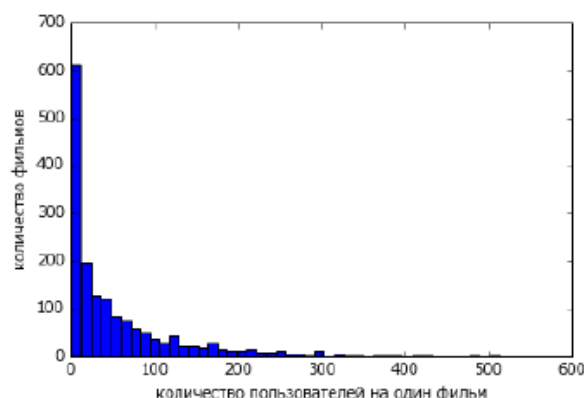


б)

Рис. 1.1 - Кількість фільмів а)позитивно б) оцінено кожним з користувачів



а)



б)

Рис. 1.2 - Кількість користувачів, а)позитивно б) оцінили кожним із фільмів

2.2 Опис підходу, що є прикладом колаборативної фільтрації

Цей підхід є прикладом колаборативної фільтрації. Введемо оператор, який у разі застосування до користувача повертає всі обрані ним об'єкти, у разі застосування до об'єкта повертає всіх користувачів, які вибрали цей об'єкт. Припустимо, дані представлені у вигляді таблиці транзакцій (Таблиця 1.1), де по стовпцях представлені об'єкти i (фільми), а по рядках користувачі u . Тоді для користувача u_1 операція повертатиме множину $\{i_1, i_2, i_4\}$, а для об'єкту i_1 повертатиме множину $\{u_1, u_3\}$.

$$u'_1 = \{i_1, i_2, i_4\}, i'_1 = \{u_1, u_3\}$$

Таблиця 1

Таблиця транзакцій

t	i_1	i_2	i_3	i_4
u_1	1	1	0	1
u_2	0	0	1	0
u_3	1	1	0	0

Побудуємо рекомендаційну систему за такими правилами. Для того, щоб зробити рекомендацію новому користувачеві u_0 , застосуємо до нього двічі операцію'. Після першого застосування отримаємо всі об'єкти, ви- лайки u_0 . Після другого зміни отримуємо всіх користувачів u_0'' , у яких серед обраних ними об'єктів трапляються всі об'єкти, вибрані u_0 .

Після цього рекомендуватимемо u_0 всі ті об'єкти, які обрали користувачі з множини u_0'' . При цьому треба не забути виключити ті об'єкти, що u_0 вже вибрав. Формально цей метод може описати так:

$$\bigcup_{u \in u_0''/u_0} u'$$

У другому прикладі спосіб знаходження об'єктів для рекомендації описується формулою.

$$(u_0''/u_0)'$$

Другий приклад, накладає жорсткіші обмеження на множину можливих рекомендацій. Пропонується знайти u_0'' , тобто. всіх користувачів, у кожного з яких безліч вибраних об'єктів містить у собі безліч об'єктів, вибраних користувачем u_0 . Далі, виключивши з отриманої множини користувача u_0 , знову застосовуємо операцію ' і отримуємо об'єкти, які вибрано всіма користувачами з множини u_0'' в сукупності, тобто, об'єкти, які були обрані всіма користувачами з u_0'' .

Зрозуміло, що в даному прикладі безліч доступних для рекомендації об'єктів ще менше, ніж у першому прикладі, зате такі рекомендації мають більший шанс виявитися правильними. Оскільки раз рекомендовані об'єкти були обрані всією групою однодумців, то швидше за все це саме те, що сподобалося б u_0 .

Третій приклад має найменш жорсткі обмеження на групу однодумців серед усіх перелічених. Для цього пропонується для кожного об'єкту обраного u_0 , знайти тих користувачів, що теж вибрали цей об'єкт, виключити з цієї множини u_0 і додати до безлічі рекомендацій ті об'єкти, які були обрані даними користувачами разом.

$$\cup_{i \in u_0'} (i'/u_0)'$$

Переваги та недоліки підходу

До переваг можна віднести відносну простоту підходу, який не потребує складних обчислювальних потужностей. Однак незважаючи на простоту та прозорість методу, він має суттєві недоліки, які виявляються при застосуванні методу до реальних даних, наприклад MovieLens. Особливість даних така, що один фільм може бути обраний малою кількістю користувачів. така частота у кожного користувача і дуже багато різних фільмів, тому коли ми шукаємо всіх користувачів, у яких серед обраних об'єктів зустрічаються ті самі, що й у u_0

					KPM.KI.0.884-03.2.3	Арк.
						26
Змн.	Арк.	№ докум.	Підпис	Дат		

тобто u_0'' зазвичай отримуємо порожню безліч. І справа тут не в недостатності даних з різних фільмів. Така особливість даних і часто всіх реальних даних, з цього випливає необхідність методів, які здатні вловити більш тонкі залежності.

2.3 Часті множини та асоціативні правила

2.3.1 Часті множини

Необхідно підрахувати частоту всіх можливих поєднань об'єктів у транзакційних даних. У нашому випадку під транзакцією розуміється набір фільмів одного користувача. Оскільки вибираються тільки фільми, позначені одиницею, тобто або фільми, що сподобалися при ненульовому порозі, і просто вибрані фільми у зворотному випадку.

Для визначення частоті множини необхідно ввести поняття мінімальної підтримки (*min support*). Є багато всіх об'єктів, то є список вибраних фільмів. Тоді підтримка множини $A \subseteq I$ визнається як кількість транзакцій, яким належить або якщо простіше, то скільки разів множина було обрано користувачами(1). При цьому враховуються й ті випадки, коли є підмножиною вибраних користувачів об'єктів.

Множина називається частою, якщо її підтримка не менше, ніж мінімальна підтримка ((2)).

$$sup(A) = \frac{\text{number of transactions containing } A}{\text{number of transactions}} \quad (2.1)$$

$$sup(A) \geq minsup \quad (2.2)$$

Перший етап полягає у знаходженні частих множин. Найгрубіший спосіб - простий перебір всіх можливих підмножин буде коштувати дуже дорого, тому що перебрати треба буде $2^{|I|}$ множин. Тому давно придумані набагато досконаліші методи. У цій роботі був використаний алгоритм FPgrowth.

					KPM.KI.0.884-03.2.3	Арк.
						27
Змн.	Арк.	№ докум.	Підпис	Дат		

Асоціативні правила

Для знаходження залежностей між наборами фільмів використовуються асоціативні правила. Нехай є множина об'єктів A і B , перетин яких порожній. Тоді асоціативне правило між A і B означає, що якщо множина A включено в транзакцію, тоді, швидше за все, множина B теж буде включено в цю транзакцію. Є дві метрики, які формалізують це "швидше за все". Перша з них, вже зазначена вище підтримка (3), тобто кількість транзакцій, що містять U , друга - достовірність (4), що показує умовну ймовірність множини при наявності множини.

$$sup(A) = \frac{\text{number of transactions containing } A \cup B}{\text{number of transactions}} \quad (2.3)$$

$$conf(A) = \frac{\text{number of transactions containing } A \cup B}{\text{number of transactions containing } A} \quad (2.4)$$

Прийнято шукати правила, підтримка та достовірність яких вищі за якісь два мінімальні значення, визначені тим, хто застосовує цей метод. Правила з високою достовірністю дають найточніші прогнози. При цьому підтримка теж важлива, тому що правила з маленькою підтримкою, тобто нечасті, не мають великої інформації і навіть можуть бути викидами.

Для того, щоб створити рекомендації, для кожного користувача потрібно знайти всі правила, такі, що користувач вибрав усі об'єкти, що знаходяться в лівій частині. При цьому це повинні бути правила, що задовольняють мінімальну підтримку та достовірність. Тоді будь-який з об'єктів, що знаходиться у правій частині правил, може бути порекомендований. Нехай R_u множина всіх об'єктів, які перебували у правій частині правил і ще не вибраних користувачем u , тоді можна відсортувати це за достовірністю правил, у яких вони були представлені, тому об'єкти з найбільшою підтримкою перебувають

					KPM.KI.0.884-03.2.3	Арк.
						28
Змн.	Арк.	№ докум.	Підпис	Дат		

вище. Якщо об'єкт був у кількох правилах, то береться максимальна достовірність. Після цього можна вибрати перші об'єкти та рекомендувати їх користувачу.

2.3.2 Алгоритм, заснований на рідкісних множинах

У секції "опис даних" виявлено те, що багато фільмів непопулярні, тобто небагато користувачів їх вибрали. Таким чином, ці об'єкти можуть бути втрачені на увазі, наприклад, при побудові частих множин. Так як якщо підтримка множини, що складається з рідкісного фільму, менша за мінімальну, то він надалі не увійде до жодної частоті множини. Тобто або треба знижувати підтримку, або вигадати якийсь додатковий спосіб навчитися створювати рекомендації, що містять ці об'єкти.

У цьому розділі пропонується алгоритм, який створює рекомендації для користувачів, які вибирають рідкісні фільми. Неформальний опис алгоритму:

1. Знайти рідкісні фільми i
2. Створити список рідкісних множин R_i , тобто. тих, які містять у собі хоча б один рідкісний фільм i
3. Для кожного користувача з тестової вибірки перетнути безліч вибраних ним до цього фільмів u'_{tr} з кожною з рідкісних множин. Для кожної рідкісної множини оцінити наскільки сильно це перетин покриває множину обраних u_{tr} об'єктів.

$$\frac{|u'_{tr} \cap R_i|}{|u'_{tr}|} \geq \varphi$$

Об'єкти з рідкісних множин, для яких нерівність виконана, додати до списку рекомендацій користувачу u .

Так як об'єктів у рекомендованій множині може виявитися більш, ніж у тестовій, та й частина з них не буде обрана, необхідно якимось чином встановити силу рекомендації, тобто. потрібно рекомендувати лише те, у чому ми впевнені. Цю впевненість доведеться встановлювати емпірично. Наприклад,

					KPM.KI.0.884-03.2.3	Арк.
						29
Змн.	Арк.	№ докум.	Підпис	Дат		

можна оцінити скільки разів кожен фільм із рекомендованих був порекомендований. Таким чином можна буде побудувати ієрархію об'єктів, і, наприклад, вибрати топ N фільмів з усіх рекомендованих.

Можна уявити двох типових користувачів, Василя і Петра. Василь дивиться переважно непопулярні фільми, наприклад, артхаус чи документальне кіно. А Петро дивиться зазвичай блокбастери і модні новинки. Таких користувачів, як Петро багато, і припустимо, у якоїсь частини з них серед вибраних фільмів зустрічаються, якісь рідкісні, зі списку Василя.

Тоді якщо потрібно буде порекомендувати щось Василеві, згідно з алгоритмом ми складемо список рекомендованих фільмів і відсортуємо його за частотою рекомендацій. Тоді може статися така ситуація, що ми обираємо, наприклад, перші 10 фільмів з цього списку, і виявляється, що це всі фільми, які мають попит у більшості таких як Петро. У результаті виявиться, що з жодним із рекомендованих Василеві фільмів ми не вгадали, тобто перетин тестової вибірки та рекомендацій порожній. Ви можете вирішити цю проблему, збільшуючи кількість рекомендованих об'єктів. Тоді є шанси порекомендувати багато непотрібного і цього постраждає точність рекомендацій та довіра Василя до наших рекомендацій. Тому одним із виходів із такої ситуації, яка дійсно зустрічалася в аналізованих даних, є виключення найпопулярніших фільмів. Оскільки даних алгоритм як створений для уловлювання рідкісних фільмів, за винятком популярних фільмів ми нічого не втрачаємо, тому що це нам може дати більшу точність у рекомендації рідкісних фільмів таким користувачам як Петро.

2.3.3 Порівняння підходів

Метрики для оцінки результатів

У цій секції належить оцінити точність запропонованого алгоритму залежно від параметрів. Є дві оцінки це точність (recision) і повнота (recall). Позначимо множину рекомендацій користувачеві u , як R , а множину обраних ним об'єктів у тестовій вибірці, як u'_t . Тоді міра точності описується за

					KPM.KI.0.884-03.2.3	Арк.
						30
Змн.	Арк.	№ докум.	Підпис	Дат		

формулою (5) та оцінює, яка частина рекомендацій була правильно передбачена. Найкращий випадок - коли все, що було рекомендовано, було обрано користувачем. Це означає, що ми пропонуємо правильні товари.

$$Precision =$$

$$Recall = \frac{|R \cap u'_t|}{|u'_t|} \quad (2.6)$$

Однак, дані оцінки конкурують між собою, оскільки можна збільшити повноту і порекомендувати безліч фільмів, що значно перевищує безліч фільмів, насправді вибраних. Таким чином, збільшується шанс вгадати з рекомендаціями, але при цьому буде порадено багато зайвого, що користувач швидше за все сприйме негативно. Так само можна і зменшити множину рекомендацій та порадити лише фільми, що відповідають найбільшій точності. Але в цьому випадку упускаються з поля зору інші, які могли б сподобатися користувачеві. Тому найкраще підходить компромісна оцінка, так само описана в [1], звана F_1 мірою (7). Тому для оцінки використовуватиметься саме F_1 міра.

$$F1 = \frac{2 \cdot recall \cdot precision}{recall + precision} \quad (2.7)$$

Порівняння параметрів

Як було зазначено в ранніх секціях, у нас є кілька параметрів, від яких залежить запропонований алгоритм і всі інші теж. Оскільки завдання полягає у складанні рекомендацій тим користувачам, які обирають рідкісні об'єкти, то й оцінювати результати необхідно саме для цих користувачів.

Тому різні параметри тим чи іншим чином впливатимуть на список цих користувачів. Відповідно, якщо навіть якийсь із порівнюваних підходів не залежить безпосередньо від цих параметрів, його результати оцінюватимуться для різних груп користувачів.

Експерименти з порогом, який визначає рідкісні об'єкти

					KPM.KI.0.884-03.2.3	Арк.
						31
Змн.	Арк.	№ докум.	Підпис	Дат		

Спочатку оцінимо, як на точність результатів впливає поріг, визначальний чи є об'єкт рідкісним чи ні. В описі алгоритму вказується, що треба знайти рідкісні фільми. Пропонується міряти, наскільки частим є фільм, за допомогою підтримки, яка вже була описана вище. Тобто вважаємо, у скільки транзакціях зустрівся фільм, і якщо це число нижче за мінімальну підтримку *minsup*, То визначаємо фільм як рідкісний. Залежність міри F_1 від цього значення представлена на графіку (3)

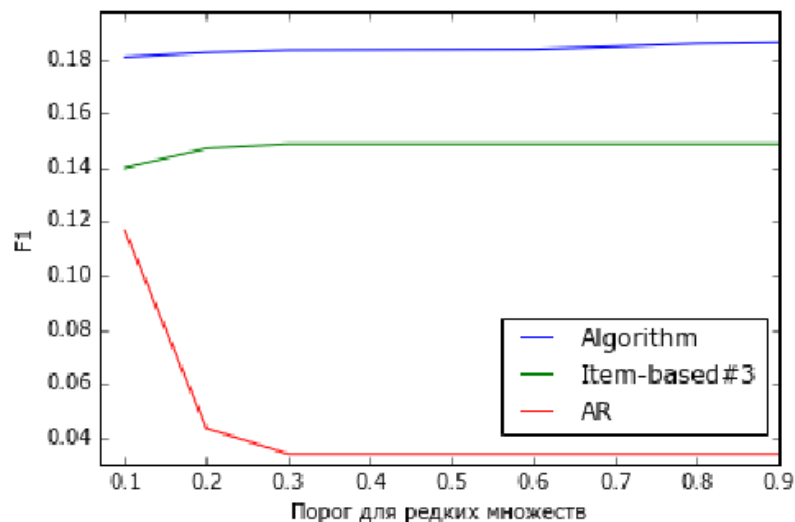


Рис. 2.3- Залежність F_1 від *minsup*

Чим вищий поріг, тим більше об'єктів відносяться до рідкісних, тим менше шансів відловити ці рідкісні фільми. Тому починаючи з порога, що дорівнює 0.3, графіки стабілізуються. Швидше за все, таке погіршення в міру F_1 *itembased* полягає в основі підходу, пов'язаного з тим, що алгоритм видає невелику кількість рекомендацій - один або два фільми. Тому, не дивлячись на добру точність, беруться до уваги інші фільми, які могли б сподобатися користувачам, тобто повнота має недостатньо великі значення.

Експерименти зі значенням

Наступний параметр, що впливає на рекомендації – це змінна ϕ , що визначає на скільки вже вибрані користувачем фільми співвідносяться з тими, що зазвичай зустрічаються з рідкісними.

На рисунку (4) показано залежність F1 від цього параметра, при цьому значення мінімального порогу вважаємо рівним 0.3.

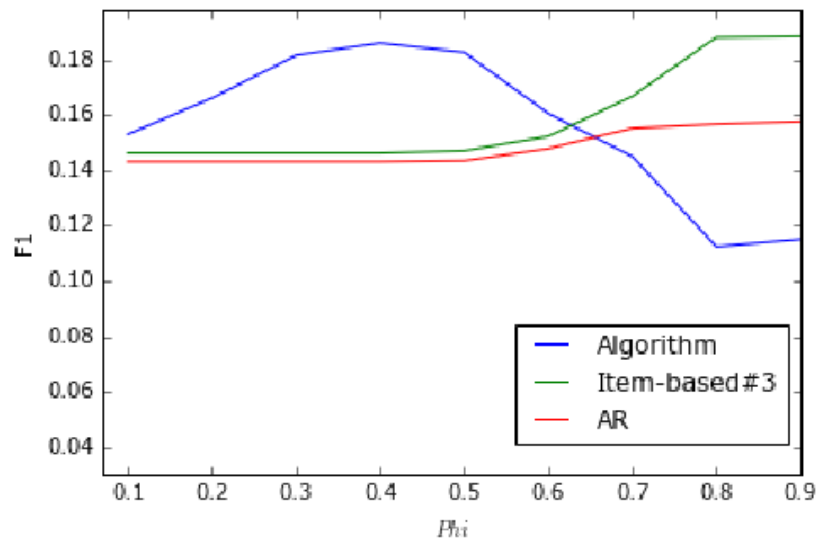


Рис. 2.4 - Залежність F1 від ϕ

З графіка видно, що базові підходи, такі як асоціативні правила та колаборативна фільтрація, зі зростанням ϕ стають точнішими. Це швидше за все пов'язано з тим, що з вибірки виключаються користувачі, які мало вибирають рідкісні фільми. Це дозволяє точніше виявляти переваги тих, кому рідкісні фільми подобаються більше.

Графік залежності запропонованого алгоритму від параметра досягає максимуму при 0.4. Можливо це обумовлено тим, що є користувачі, у яких безліч обраних товарів містить у собі багато рідкісних множин, такий випадок задовольняє більшим ϕ . Тобто нам попався користувач, котрий уже вибрав багато фільмів і тому алгоритму складно знайти, що саме йому рекомендувати. Точність для такого користувача невелика для всіх значень ϕ , але при цьому для $\phi \geq 0.6$ це стає критичним. То є виявляється те, що алгоритм гірше справляється з тими, у кого вже багато фільмів вибрано.

Експерименти з кількістю рекомендованих об'єктів

					KPM.KI.0.884-03.2.3	Арк.
						33
Змн.	Арк.	№ докум.	Підпис	Дат		

Важливо також оцінити скільки значень ми хочемо порекомендувати користувачеві. Один із варіантів, коли фільмів небагато, зате вони точно будуть переглянуті. З іншого боку користувач може сам не знати чого він хоче, тоді слід запропонувати йому більший список і що-небудь з нього він точно вибере. Тому на рисунку (5) представлено те, як наша міра точності змінюється залежно від кількості рекомендованих фільмів.

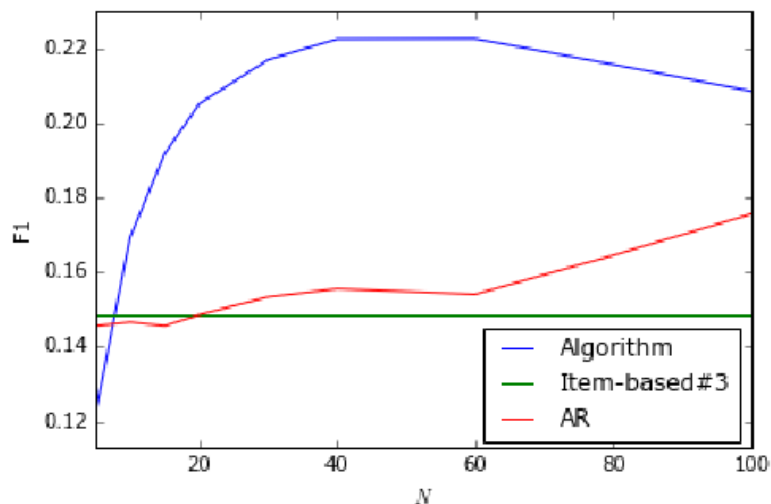


Рис. 2.5 - Залежність F1 від Φ

З графіка видно, що найкраще всі підходи працюють, якщо на до рекомендувати близько 50 фільмів, хоча вже на 15 запропонований алгоритм дає досить добрий результат.

Експерименти з порогом для оцінки фільму

Раніше вже було сказано, що спочатку дані були оцінки, поставлені користувачами різним фільмам, і в нашій владі було те, як їх трактувати. До цього всі експерименти були для випадку, коли оцінка не менше 3 - фільм сподобався, значить ставимо 1, нижче 3 - ставимо 0 і розуміємо, що фільм не сподобався. У цьому підрозділі дізнаємося, що буде, якщо всі інші параметри залишити незмінними, але поміняти поріг з 3 на 0. Тобто прогнозуватимемо, чи буде фільм взагалі оцінений, тому що ми припускаємо, що фільм спочатку потрібно подивитися, а потім оцінювати, то можна це трактувати, як буде фільм обраний. У таблиці (2) наведено, як зміниться точність, якщо ми знизимо поріг.

Залежність F1 міри від порогу шкалювання

<i>Algorithm</i>	cutoff = 3	cutoff = 0
AR	0.15	0.14
IB3	0.16	0.15
Algorithm	0.21	0.18

Висновки

У цій роботі було порівняно 3 способи побудов рекомендаційної системи. Перший ґрунтувався на колаборативній фільтрації, інший на асоціативних правилах і останнім був алгоритм, запропонований у цій роботі. Основним завданням було дізнатися, як поведуться ці алгоритми у випадку з користувачами, які обирають рідкісні фільми.

Було проведено низку випробувань, які встановлювали вплив різних параметрів на значення I заходи, що оцінює сукупність точності та повноти. Було оцінено такі параметри, як мінімальна підтримка, яка визначала, чи фільм є рідкісним, поріг, який визначав, які користувачі надають перевагу рідкісним фільмам, кількості рекомендованих фільмів і поріг для рейтингів, який відповідав за зміст рекомендацій, тобто це фільми, які повинні сподобатися, або прогнозувався факт перегляду і не враховувалася задоволеність фільмом.

РОЗДІЛ 3

РЕАЛІЗАЦІЯ МЕТОДУ, РОЗРОБКА ПРИКЛАДУ СИСТЕМИ ВІДБОРУ ВІДЕОМАТЕРІАЛІВ НА ОСНОВІ АСОЦІАТИВНИХ ПРАВИЛ

2.1 Оцінка якості та тестування рекомендаційної системи

Тестування рекомендаційної системи - процес непростий і завжди викликає безліч питань, головним чином через неоднозначність поняття «якість».

Загалом у завданнях машинного навчання є два основні підходи до тестування: offline тестування моделі на історичних даних за допомогою ретротестів, тестування готової моделі за допомогою A/B тестування (запускаємо кілька варіантів, дивимось який дає найкращий результат).

Обидва ці підходи активно застосовуються розробки рекомендаційних систем. Почнемо з offline тестування.

Основне обмеження, з яким доводиться стикатися, — оцінити точність прогнозу ми можемо лише на тих товарах, які вже оцінив користувач.

Стандартний підхід – це крос-валідація методами leave-one-out та leave-p-out. Багаторазове повторення тесту із усередненням результатів дозволяє отримати більш стійку оцінку якості.

leave-one-out - модель навчається на всіх оцінених користувачем об'єктах, крім одного, а тестується на цьому об'єкті. Так робиться всім n об'єктів, і серед отриманих n оцінок якості обчислюється середнє.

leave-p-out - те саме, але на кожному кроці виключається p точок.

Усі метрики якості можна умовно поділити на три категорії:

Prediction Accuracy - оцінюють точність прогнозованого рейтингу,

Decision support - оцінюють релевантність рекомендацій,

Rank Accuracy метрики - оцінюють якість ранжування рекомендацій, що видаються.

					KPM.KI.0.884-03.2.3	Арк.
						36
Змн.	Арк.	№ докум.	Підпис	Дат		

На жаль, не існує єдиної рекомендованої метрики на всі випадки життя і кожен, хто займається тестуванням рекомендаційної системи, підбирає її під цілі.

Коли рейтинги оцінюються за безперервною шкалою (0-10), зазвичай досить метрик класу Prediction Accuracy.

Таблиця 3.1

Метрики класу Prediction Accuracy

Назва	Формула	Опис
MAE (Mean Absolute Error)	$E(P - R)$	Середнє абсолютне відхилення
MSE (Mean Squared Error)	$E(P - R ^2)$	Середньоквадратична помилка
RMSE (Root Mean Squared Error)	$\sqrt{E(P - R ^2)}$	Корінь з середньоквадратичної помилки

Метрики класу Decision Support працюють з бінарними даними (0 та 1, та й ні). Якщо в нашому завданні рейтинги спочатку відкладаються за безперервною шкалою, їх можна перевести в бінарний формат, застосувавши вирішальне правило — скажімо, якщо оцінка менша за 3.5, вважаємо оцінку «поганою», а якщо більшою, то «хорошою».

Таблиця 3.2

Метрики класу Decision Support

Назва	Формула	Опис
Precision	$\frac{TP}{TP + FP}$	Частка рекомендацій, що сподобалися користувачеві
Recall	$\frac{TP}{TP + FN}$	Частка цікавих для користувача товарів, яка показана

F1-Measure	$\frac{2PR}{P + R}$	Середній гармонійний метрик Precision і Recall. Корисно, коли заздалегідь неможливо сказати, яка з метрик важливіша
ROC AUC		Наскільки високою є концентрація цікавих товарів на початку списку рекомендацій
Precision@N		Метрика Precision, порахована на Top-N записах
Recall@N		Метрика Recall, порахована на Top-N записах
AverageP		Середнє значення Precision по всьому списку рекомендацій

Зазвичай, рекомендації виводяться списком із кількох позицій (спочатку топова, потім у порядку зменшення пріоритету). Метрики класу Rank Accuracy вимірюють наскільки правильним є порядок показу рекомендацій у відсортованому списку.

Якщо ми візьмемо рекомендаційні системи в онлайн бізнесі, то вони, як правило, мають дві (іноді суперечливі) цілі: проінформувати користувача про цікавий товар, спонукати його здійснити покупку (шляхом розсилки, складання персональної пропозиції тощо).

					KPM.KI.0.884-03.2.3	Арк.
						38
Змн.	Арк.	№ докум.	Підпис	Дат		

Метрики класу Rank Accuracy

Назва	Формула	Опис
Mean Reciprocal Rank	$E\left(\frac{1}{pos}\right)$	На якій позиції списку рекомендацій користувач знаходить першу корисну
Spearman Correlation	$E(P - R ^2)$	Кореляція (Спірмена) реального та прогнозованого рангів рекомендацій
nDCG	$\sum \frac{R(i)}{\max(1, \log(i))}$	Інформативність видачі з урахуванням ранжування рекомендацій
Fraction of Concordance Pairs	$P(X_R > X_P)$	Наскільки високою є концентрація цікавих товарів на початку списку рекомендацій

Як і в будь-якій моделі, спрямованій на мотивацію користувача до дії, слід оцінювати тільки інкрементальний приріст цільової дії. Тобто, наприклад, при підрахунку покупок за рекомендацією нам потрібно виключити ті, які користувач і так зробив би без нашої моделі. Якщо цього не зробити, ефект від застосування моделі буде сильно завищений.

Lift — показник того, у скільки разів точність моделі перевершує якийсь базовий алгоритм. У нашому випадку базовий line алгоритм може бути просто відсутність рекомендацій. Дана метрика добре відловлює частку інкрементальних покупок, і це дозволяє ефективно порівнювати різні моделі.

					KPM.KI.0.884-03.2.3	Арк.
						39
Змн.	Арк.	№ докум.	Підпис	Дат		

Поведінка користувача рідко погано формалізується і жодна метрика повною мірою не опише розумові процеси у його голові під час виборів товару. На рішення впливає багато чинників. Перехід по посиланню з товаром, що рекомендується, ще не є його високою оцінкою або навіть інтересом. Частково зрозуміти логіку клієнта допомагає онлайн-тестування. Нижче наводиться пара сценаріїв такого тестування.

Перший і найочевидніший сценарій – аналіз подій сайту. Ми дивимось, що користувач робить на сайті, чи звертає увагу на наші рекомендації, чи переходить за ними, які фічі системи мають попит, які — ні, які товари краще рекомендуються, які гірші. Щоб зрозуміти який із алгоритмів загалом працює краще або просто спробувати нову перспективну ідею, робимо А/В тестування та збираємо результат.

Другий сценарій – це отримання фідбеку від користувачів у вигляді опитувань та голосувань. Як правило, це загальні питання для розуміння, як клієнти користуються сервісом — що важливіше: релевантність або різноманітність, чи можна показувати продукти, що повторюються, або це занадто дратує. Перевага сценарію – він дає пряму відповідь на всі ці питання.

Подібне тестування штука складна, але для великих рекомендацій вона просто необхідна. Питання можуть бути і складнішими, наприклад, «який зі списків вам здається більш релевантним», «наскільки лист виглядає цілісним», «Чи будете дивитися цей фільм/читати книгу».

3.2 Розробка прикладу рекомендаційної системи відбору відеоматеріалів на основі асоціативних правил

3.2.1 Етапи створення рекомендаційної системи на основі асоціативних правил

Створення рекомендаційних систем на основі асоціативних правил – це цікавий підхід, який ґрунтується на виявленні кореляцій між різними

					<i>KPM.KI.0.884-03.2.3</i>	Арк.
						40
Змн.	Арк.	№ докум.	Підпис	Дат		

елементами чи товаровими позиціями. Ось кілька кроків для розробки такої системи:

Крок 1: Збір даних

Зібрати дані про покупки або інші взаємодії користувачів з продуктами/сервісами. Це може бути інформація про покупки в інтернет-магазині, перегляди відео, кліки на веб-сайті тощо.

Крок 2: Підготовка даних

Перетворення даних у формат, який підходить для аналізу асоціативними правилами. Найчастіше це виглядає як таблиця, де кожен рядок представляє унікальну транзакцію, а кожен стовпчик відповідає конкретному продукту або елементу, що брав участь у транзакції. Це може бути виконано за допомогою one-hot encoding або інших методів.

Крок 3: Виявлення частих наборів (itemsets)

Використання алгоритмів, таких як Apriori або FP-росту, для виявлення часто зустрічаються наборів продуктів, які часто купуються разом.

Крок 4: Визначення асоціативних правил

На основі частих наборів створення асоціативних правил. Правила зазвичай мають формат "Якщо користувач купує А, то ймовірно купить В".

Крок 5: Оцінка правил

Оцінка створених асоціативних правил за допомогою метрик, таких як підтримка (support), достовірність (confidence), підтримка або інші метрики, що вказують на силу та значущість правила.

Крок 6: Рекомендації

Використання створених правил для надання рекомендацій користувачам. Наприклад, якщо користувач переглядає товар А, система може рекомендувати товар В на основі асоціативного правила.

Крок 7: Тестування та покращення

Тестування системи на нових даних, виявлення сильних та слабких сторін системи та вдосконалення алгоритмів на основі отриманих результатів.

					KPM.KI.0.884-03.2.3	Арк.
						41
Змн.	Арк.	№ докум.	Підпис	Дат		

За допомогою цих кроків можна створити просту рекомендаційну систему на основі асоціативних правил. У реальних застосуваннях часто використовуються більш складні алгоритми та методи для покращення точності та рівня персоналізації рекомендацій.

Зважаючи на те, що рекомендаційні системи на основі асоціативних правил реалізуються на основі виявлення асоціацій між різними елементами, ось спрощений приклад на Python, який покаже, як можна використати бібліотеку mlxtend для виявлення асоціативних правил на основі мінімального набору підтримки та достовірності.

Для початку встановіть бібліотеку mlxtend:

```
bash
```

```
Copy code
```

```
pip install mlxtend
```

Ось приклад коду для виявлення асоціативних правил на основі мінімального набору підтримки та достовірності:

```
python
```

```
Copy code
```

```
from mlxtend.preprocessing import TransactionEncoder
from mlxtend.frequent_patterns import apriori, association_rules

# Приклад даних (список списків з покупками)
dataset = [
    ['яблуко', 'пиво', 'молоко'],
    ['яблуко', 'пиво'],
    ['яблуко', 'молоко'],
    ['яблуко', 'пиво']
]

# Перетворення даних у формат, який вимагає mlxtend
```

					KPM.KI.0.884-03.2.3	Арк.
						42
Змн.	Арк.	№ докум.	Підпис	Дат		

```

te = TransactionEncoder()
te_ary = te.fit(dataset).transform(dataset)
df = pd.DataFrame(te_ary, columns=te.columns_)

# Застосування Apriori для виявлення часто зустрічаються наборів
елементів

frequent_itemsets = apriori(df, min_support=0.5, use_colnames=True)

# Виявлення асоціативних правил

rules = association_rules(frequent_itemsets, metric="confidence",
min_threshold=0.7)

# Виведення асоціативних правил

print(rules)

```

У цьому прикладі дані представлені у форматі списку списків (dataset), де кожен список представляє одну транзакцію або покупку. Потім використовується бібліотека mlxtend для виявлення часто зустрічаються наборів елементів (frequent_itemsets) та створення асоціативних правил (rules) з використанням метрик, таких як підтримка та достовірність.

Цей код є досить загальним прикладом роботи з алгоритмами виявлення асоціативних правил у Python. Реальна реалізація рекомендаційної системи на основі асоціативних правил може включати додаткові кроки, такі як обробка даних, оптимізація параметрів алгоритмів та інші етапи розробки.

Принцип роботи асоціативних правил базується на виявленні зв'язків між елементами в наборі даних. Головна мета - виявити часто зустрічаються комбінації елементів (товарів, подій, характеристик тощо), які з'являються разом.

Основні кроки роботи асоціативних правил:

1. Підготовка даних

Набір даних має бути у вигляді таблиці, де кожен рядок представляє окрему транзакцію, а кожен стовпчик відповідає наявності чи відсутності певного елемента в цій транзакції.

2. Визначення підтримки

					KPM.KI.0.884-03.2.3	Арк.
						43
Змн.	Арк.	№ докум.	Підпис	Дат		

Підтримка (support) вказує на те, яка частка транзакцій містить певну комбінацію елементів. Наприклад, якщо підтримка елементів; хліб; і ;молоко ; дорівнює 0.2, це означає, що ці два товари купують разом у 20% від усіх транзакцій.

3. Визначення достовірності

Достовірність (confidence) вказує на ймовірність того, що після покупки одного товару буде придбаний інший. Наприклад, якщо достовірність для правила ;хліб ; молоко; дорівнює 0.8, це означає, що у 80% випадків, коли купувався хліб, також купувалось молоко.

4. Відбір асоціативних правил

Алгоритми, такі як Apriori чи FP-дерева, використовуються для виявлення часто зустрічаються зв'язків між елементами на основі встановлених значень підтримки та достовірності.

5. Генерація рекомендацій

На основі виявлених асоціативних правил можна генерувати рекомендації.

Наприклад, якщо користувач купує товар А, система може рекомендувати йому товар В на основі асоціативного правила А -В;.

Асоціативні правила допомагають виявляти зв'язки між елементами, що може бути корисним для різних сфер, включаючи рекомендації покупок, аналіз бізнес-процесів, маркетингові стратегії тощо.

3.2.1 Опис прикладу відбору відеоматеріалів

Завдання знаходження асоціативних правил розбивається на дві підзадачі:

- Знаходження всіх наборів елементів (таких, що часто зустрічаються), які задовольняють порогу minsupport.
- Генерація правил з наборів елементів, що знайдені згідно з достовірністю, яка задовольняє порогу minconfidence.

					KPM.KI.0.884-03.2.3	Арк.
						44
Змн.	Арк.	№ докум.	Підпис	Дат		

Алгоритм Apriori має просту структуру, завдяки якій в неї можна легко вносити модифікації для оптимізації алгоритму під конкретні набори даних для вирішення задачі пошуку асоціативних правил.

У якості використаного алгоритму обрано алгоритм Apriori. Цей алгоритм базується на методах інтелектуального аналізу (Data Mining), призначений для виявлення знань в базах даних. Пошук асоціативних правил - ключова тема у інтелектуальному аналізі. В результаті можливий пошук приховані зв'язки в, на перший погляд, ніяк не пов'язаних даних. Ці зв'язки - асоціативні правила. Ті правила, кількість (підтримка) яких перевищує певний поріг, вважаються цікавими. Прикладом такого правила, є твердження, що в тому випадку, якщо відбулася подія А, то станеться і подія В з імовірністю Х. Одним з найбільш часто цитованих прикладів пошуку асоціативних правил служить проблема пошуку стійких зв'язків в кошику покупця (англ. Market Basket Problem). Проблема пошуку стійких зв'язків в кошику покупця полягає в тому, щоб визначити які товари купуються покупцями разом, так, щоб фахівці з маркетингу могли відповідним чином розмістити ці товари в магазині для підвищення обсягу продажів. Деякі виявляються правила можуть бути тривіальними, наприклад: «покупці, які купують пиво, також купують і горішки». Інші - цікаві і не тривіальні, наприклад: «покупці, які купують мило, так само купують і пиво». Саме здатність виявляти цікаві правила робить пошук асоціативних правил цінним і сприяє пошуку знань. Також як приклад завдань пошуку асоціативних правил можна розглянути безлічі діагнозів, що зустрічаються одночасно у одного хворого, набори факторів ризику, що призводять до аварій на виробництві і т.д.

Асоціативні правила зазвичай мають дві частини: антецедент (припущення або умова) та консеквент (наслідок). У попередньому прикладі "покупець купує хліб" - це антецедент, а "з ймовірністю 70% він також купить молоко" - консеквент.

Для визначення асоціативних правил часто використовують алгоритми, такі як Apriori або FP-дерева, які аналізують набори даних для виявлення часто

					KPM.KI.0.884-03.2.3	Арк.
						45
Змн.	Арк.	№ докум.	Підпис	Дат		

зустрічаються зв'язків між елементами. Такі правила можуть бути корисними для рекомендаційних систем, стратегій маркетингу, оптимізації розміщення товарів у магазинах та інших сферах, де важливо розуміти зв'язки між різними подіями чи елементами.

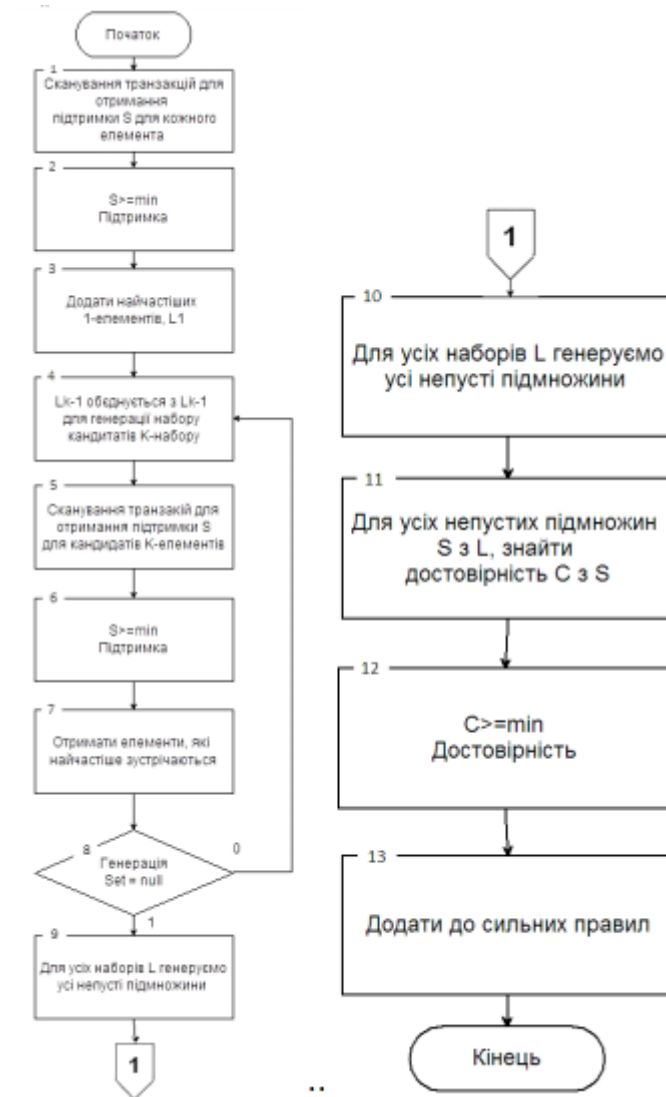


Рис 3.1 - Низькорівнева схема алгоритму

Для програмної реалізації системи обрано мови програмування JavaScript та TypeScript.

Було здійснено порівняння, найбільш популярних REST API Node.js фреймворків, яке показало, що за переліком переваг найбільш доцільно використовувати Loopback.

Обґрунтовано вибір технології реалізації програмних засобів клієнтської частини.

Здійснено порівняльний аналіз трьох найбільш популярних клієнтських фреймворків.

Для розробки клієнтської частини було обрано Ang

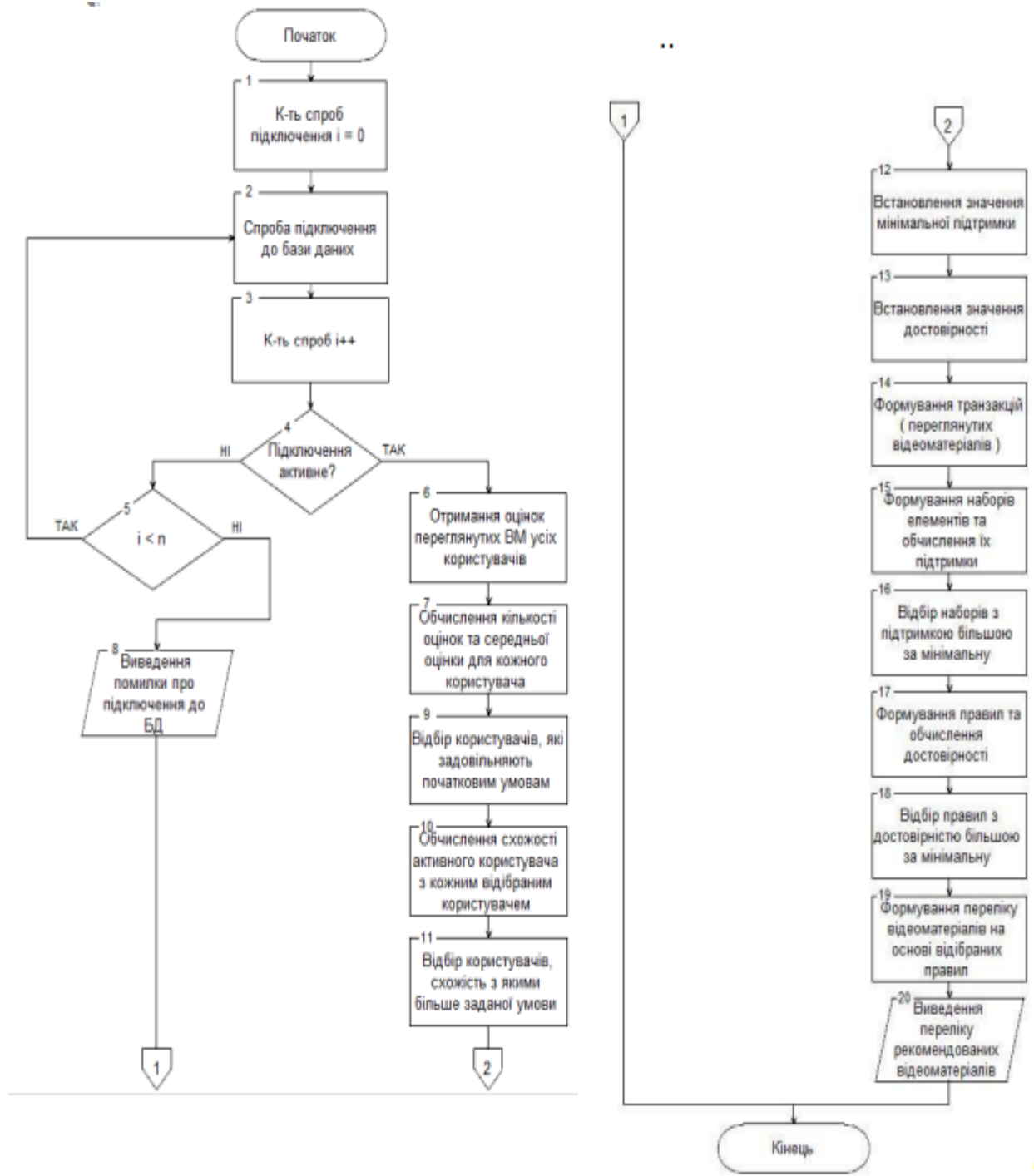


Рис 3.2 - Алгоритм роботи комбінованих алгоритмів

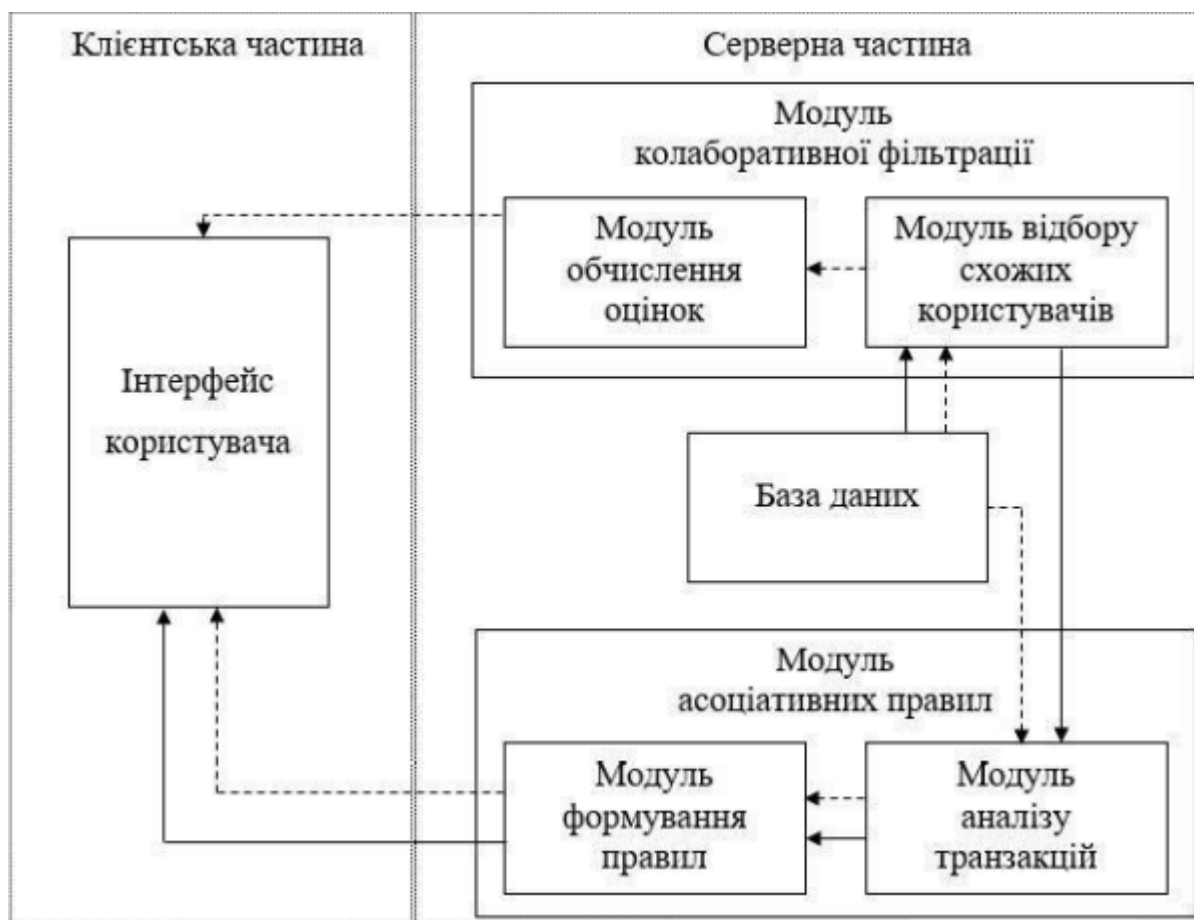


Рис. 3.3 - Загальна структурна схема інформаційної технології відбору відеоматеріалів

Тестування проводилося на обліковому записі jason_statham@gmail.com, якому було присвоєно перегляд 100 фільмів. Вподобання та оцінки фільмів є суб'єктивними.

Алгоритм розпочав своє функціонування після 10 заповнених облікових записів. До цього дані змінювалися, але це було не суттєво.

Із збільшенням числа переглянутих фільмів рекомендації поступово змінювалися і починали відповідати вподобанням. Отже виявлено перший недолік системи, а саме високий поріг входу. Користуючись даним алгоритмом користувач не зможе зразу оцінити переваги системи.

Після введення більше 100 фільмів дані алгоритми розпочали рекомендацію дійсно цікавих фільмів на погляд користувача.

Тестування швидкодії процесу надання рекомендацій Здійснювалося на аккаунті користувача `jason_statham@gmail.com`, який оцінив 100 фільмів. В таблиці 3.5 наведено схожість даного користувача до інших. Поріг схожості складає 0.3 На таблиці зображено схожість користувача `jason_statham` з іншими користувачами

Таблиця 3.4

Схожість користувача `jason_statham` з іншими користувачами

№	UserID	Схожість
1	1	0.3316156231040968
2	2	0.36270665303380223
3	4	0.3541802926436001
4	5	0.35591542502953116
5	6	0.4003608339284706
6	7	0.3090738643394098
7	10	0.31469648083921886

Тестові умови:

Колаборативна фільтрація рекомендує 8 відеоматеріалів.

Асоціативні правила – 5.

Комбінований алгоритм – 5.

В даному випадку комбінований алгоритм та асоціативні правила генерують однакові фільми, що дозволяє достовірно оцінити час роботи кожного з алгоритмів та порівняти швидкодію.

Процесор: Intel Core i3-6100 3.7GHz/8GT/s/3MB

Материнська плата : MSI H110M PRO-VH

ОЗУ : Kingston DDR4-2133 8192MB

SSD: Kingston SSDNow V300 120GB

Node.js : v8.8.1

Середнє значення роботи алгоритму Apriori складає 6488.1 мс., а комбінованого – 370.9 мс., що в 17,493 рази швидше.

					KPM.KI.0.884-03.2.3	Арк.
						49
Змн.	Арк.	№ докум.	Підпис	Дат		

Час виконання алгоритмів

№	Назва алгоритму	Час виконання (мс)
1	Apriori	6563
	Комбінований	400
2	Apriori	6707
	Комбінований	367
3	Apriori	6030
	Комбінований	346
4	Apriori	6384
	Комбінований	362
5	Apriori	6455
	Комбінований	350
6	Apriori	6432
	Комбінований	368
7	Apriori	7384
	Комбінований	365
8	Apriori	6114
	Комбінований	354
9	Apriori	6242
	Комбінований	432
10	Apriori	6570
	Комбінований	365

Наукова новизна одержаних результатів.

Підвищено швидкодію процесу формування асоціативних правил за рахунок впровадження попередньої фільтрації транзакцій. Запропоновано інформаційну технологію відбору відеоматеріалів на основі асоціативних правил та колаборативної фільтрації, що забезпечило підвищення швидкодії та якості надання рекомендації.

Запропонована інформаційна технологія сприяє підвищенню ефективності процесу надання рекомендації, зокрема: Розроблено алгоритм для підвищення ефективності надання рекомендацій на основі комбінованого застосування алгоритмів асоціативних правил та колаборативної фільтрації; Реалізовано програмний засіб на сучасних технологіях розробки: Loorback 3.0, Fireloop, Angular4, що забезпечує підвищення надійності та швидкодії. Розроблений програмний засіб для підвищення продуктивності надання рекомендацій, що дозволяє користувачам застосовувати комбіновані алгоритми

асоціативних правил та колаборативної фільтрації, зберігати їхню історію переглядів і залишати оцінки відеоматеріалам.

Висновки до третього розділу

Описано процес оцінки та тестування рекомендаційних систем.

Описано етапи створення рекомендаційної системи на основі асоціативних правил.

Розроблено та проаналізовано приклад роботи рекомендаційної системи відбору відеоматеріалів на основі асоціативних правил та колаборативної фільтрації, що забезпечило підвищення швидкодії та якості надання рекомендації.

Запропонована інформаційна технологія сприяє підвищенню ефективності процесу надання рекомендації, зокрема: Розроблено алгоритм для підвищення ефективності надання рекомендацій на основі комбінованого застосування алгоритмів асоціативних правил та колаборативної фільтрації.

					KPM.KI.0.884-03.2.3	Арк.
						51
Змн.	Арк.	№ докум.	Підпис	Дат		

РОЗДІЛ 4

ЕКОНОМІЧНА ЧАСТИНА

4.1 Техніко-економічний аналіз передбачуваного проекту

4.1.1 Класифікаційна оцінка проекту

Клас – монопроект, так як це проект з чіткими завданнями.

Тип – організаційно-технічний: забезпечення якості обслуговування користувачів мережі доступу.

Вид – дослідно-інноваційний.

Тривалість – короткостроковий.

Рівень – галузевий.

4.1.2 Призначення проекту

4.1.3 Життєвий цикл проекту

Поняття життєвого циклу проекту важливе для дослідження й аналізу проблем фінансування пов'язаних з ним робіт і прийняття відповідних управлінських рішень під час його реалізації.

Життєвий цикл проекту (проектний цикл) – це період від народження ідеї до завершення та закриття проекту.

Проектний цикл слід віднести окремі стадії або так звані етапи розвитку, які називаються фазами. Найчастіше виокремлюють саме: доінвестиційну, реалізації та експлуатації.

До інвестиційна фаза включає аналіз умов для втілення проектного задуму; розробку концепції проекту; розробку бізнес-плану та попереднє обґрунтування інвестицій, оцінку життєздатності проекту.

На етапі розробки концепції проекту визначають кінцеві цілі проекту і виявляють можливі шляхи їх досягнення.

Цей етап охоплює в собі чітко сформульовані основні характеристики проекту, до яких в свою чергу відносяться:

					KPM.KI.0.884-03.2.3	Арк.
						52
Змн.	Арк.	№ докум.	Підпис	Дат		

- наявність альтернативних технічних і технологічних можливостей;
- тривалість та попит на продукцію проекту;
- рівень базових і прогнозованих цін на послуги;
- перспективи експорту продукції;
- складність проекту;
- співвідношення витрат на реалізацію проекту і його результатів.

На основі цих та інших показників попередньо аналізують можливості реалізації проекту, нерідко для таких цілей використовується допомогою експертна система. Етап оцінки життєздатності проекту передбачає стисле ТЕО. Далі формують конкретні цілі й обмеження, а також визначають вартість проекту з точністю 25 – 40%.

Результатом такої оцінки життєздатності проекту є обґрунтування переваги обраної альтернативи перед іншими. Після цього інвестор, або замовник має переконатися, що вироблена, в результаті реалізації проекту, продукція, протягом життєвого циклу, матиме стабільний попит, достатній для призначення ціни, яка б забезпечила покриття витрат на експлуатацію й обслуговування об'єктів проекту, швидку окупність капіталовкладень.

Проектний продукт має бути підданий таким видам аналізу: технічному, комерційному, екологічному, організаційному, соціальному.

Розглянемо ті критеріями життєздатності проекту, які класифіковані на початку оцінки проекту.

Технічний аналіз дає змогу виявити техніко-економічні альтернативи; варіанти місцезнаходження об'єкта; масштаб і обсяг проекту; терміни реалізації проекту загалом і за фазами; доступність і достатність сировини та інших необхідних ресурсів.

Організаційний – оцінка організаційних, правових, політичних та адміністративних умов, в яких має реалізуватися й експлуатуватися проект.

Після визначення життєздатності проекту і прийняття рішення про початок його здійснення складають план робіт, тобто структурно визначену

					KPM.KI.0.884-03.2.3	Арк.
						53
Змн.	Арк.	№ докум.	Підпис	Дат		

послідовність етапів робіт, які виконують задля досягнення визначеного комплексу цілей (хто й що має робити і в які терміни). На основі плану робіт складають докладний календарний графік робіт або мережного графіку, що дає змогу точніше оцінити вартість проекту.

4.2 Організаційне забезпечення проекту

4.2.1. До основних завдань, вирішення яких веде до досягнення поставленої мети належать наступні.

Таблиця 4.1 – Склад робіт проекту, їх тривалість

Ко д ро боти	Найменування роботи	Т (дні)	Код Попере дньої роботи
0-1	Розробка технічного завдання	9	-
1-2	Збір даних	7	0-1
2-3	Аналіз існуючої мережі	7	1-2
3-4	Проведення аналізу положення	9	2-3
4-5	Встановлення потреби в результатах	5	3-4
4-6	Затвердження концепції	4	3-4
5-7	Встановлення ділових контактів, вивчення цілей, мотивів, вимог	5	4-5
6-7	Розвиток концепції, планування наочної області інших елементів проекту	4	4-6
7-	Розробка і затвердження	6	4-7; 6-7

					KPM.KI.0.884-03.2.3	Арк.
						54
Змн.	Арк.	№ докум.	Підпис	Дат		

8	звідного плану		
7-9	Організація виконання робіт	10	4-7; 6-7
8-10	Інформаційний контроль за виконанням робіт	6	7-8
9-10	Детальне проектування і технічні специфікації	8	7-9
10-11	Керівництво і координація робіт, коректування основних показників проекту	10	8-10; 9-10
11-12	Підтвердження закінчення робіт	4	10-11
11-13	Експлуатаційні випробування остаточного продукту проекту	6	10-11
12-14	Підготовка кадрів по експлуатації проекту	5	11-12
13-14	Підготовка документів і здача проекту замовникові	9	11-13
14-15	Оцінка результатів проекту і підведення підсумків	4	12-14; 13-14
15-16	Оцінка підсумкових документів і закриття проекту	10	14-15
		$\Sigma=1$ 10	

4.2.3 Побудова мережного графіка

За даними таблиці 4.1 будується мережевий графік проекту:

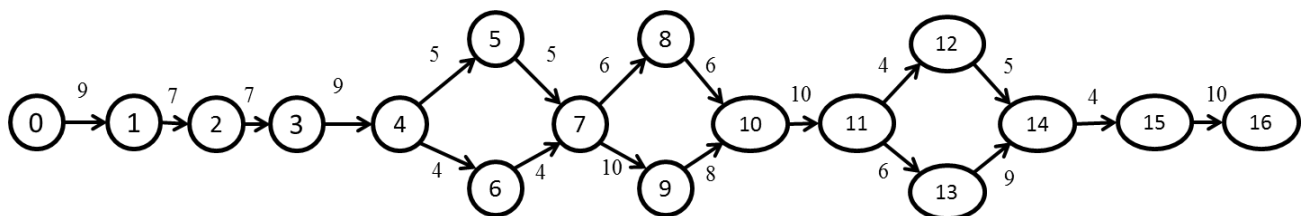


Рис. 4.1 - Мережевий графік проекту

4.2.4 Розрахунок параметрів робіт мережного графіка

					KPM.KI.0.884-03.2.3	Арк.
						55
Змн.	Арк.	№ докум.	Підпис	Дат		

Проводиться розрахунок параметрів мережного графіка:

- $t_{p.n.}$ – ранній початок роботи;
- $t_{p.o.}$ – раннє закінчення роботи;
- $t_{п.н.}$ – пізніший початок роботи;
- $t_{п.о.}$ – пізнє закінчення роботи;
- R_c – вільний резерв часу;
- R_i – резерв часу події;
- Lk_p – тривалість критичного шляху.

Розрахунок параметрів мережного графіку зведемо у таблицю 4.2:

Таблиця 4.2

Параметри робіт мережного графіка

Поп ередня робот а	Робота		Ранній строк			Пізній строк			R_i
	i	j	$t_{pн}$	t_{ij}	t_{po}	$t_{пн}$	$t_{по}$	R_c	
-	0	1	1	9	9	2	10	100	0
0-1	1	2	9	7	15	10	16	94	1
1-2	2	3	15	7	21	16	22	88	1
2-3	3	4	21	9	29	22	30	80	1
3-4	4	5	29	5	33	30	34	76	0
3-4	4	6	33	4	36	34	37	73	1
4-5	5	7	36	5	40	37	41	69	1
4-6	6	7	40	4	43	41	44	66	1
4-7; 6-7	7	8	43	6	48	44	49	61	1
4-7; 6-7	7	9	48	10	57	49	58	52	1
7-8	8	10	57	6	62	57	63	47	1
7-9	9	10	62	8	69	63	70	40	0
8-10; 9-10	10	11	69	10	78	70	79	31	1
10-11	11	2	8	4	83	79	84	26	1
10-11	11	13	81	6	86	82	87	23	0
11-12	12	14	86	5	90	87	91	19	0

11-13	13	14	90	9	98	91	99	11	0
12-14; 13-14	14	15	98	4	101	99	102	8	1
14-15	15	16	101	10	110	102	111	0	1

Визначення критичного шляху: 0-1-2-3-4-6-7-8-10-11-12-14-15-16

4.2.5 Маркетингове обґрунтування проекту

Даний проект є актуальним, особливо в зв'язку з подіями останнього часу, коли збільшується кількість користувачів, а відповідно і навантаження на мережу. Провівши аналіз – зрозуміло, що на необхідність модернізації найбільше впливали наступні фактори. Поява нових користувачів – молодшого віку (школярі молодших класів), які раніше не користувалися інтерактивними послугами під час навчання і старшого віку (вчителі старшого віку), які раніше взагалі не користувалися онлайн послугами, а зараз змушені освоювати комп'ютери та виходити в онлайн для роботи. Необхідність в нових ІКП послугах інтерактивного характеру, саме для них важлива якість (мінімізація помилок, час доступу та ін.) роботи. Все фактори, які зазначені вище збільшили навантаження на мережу.

До чинників, що визначають доцільність впровадження проекту з модернізацією мережі відносяться:

- підвищення оперативності роботи;
- підвищення надійності ходу технологічного процесу;
- зниження до мінімального значення кількості помилок;
- вивільнення персоналу до оптимальної кількості.

4.3 Економічні розрахунки проекту

4.3.1 Визначення трудомісткості розробки ПП

					KPM.KI.0.884-03.2.3	Арк.
						57
Змн.	Арк.	№ докум.	Підпис	Дат		

Розрахунок трудомісткості програмного продукту, що розробляється, проводиться за формулою:

$$T_{ПП} = T_{ТЗ} + T_{ТП} + T_{РП} + T_{ВН}, \quad (4.1)$$

де: $T_{ТЗ}$ – трудомісткості розробки технічного завдання створення ПП;

$T_{ТП}$ – трудомісткості розробки технічного проекту ПП;

$T_{РП}$ – трудомісткості розробки робочого проекту ПП;

$T_{ВН}$ – трудомісткості впровадження розробленого ПП;

Розраховуючи трудомісткість, необхідно враховувати, що програмний засіб (ПЗ) по ступеню новизни, що розробляється, є ПЗ, яке має аналог. Крім того, по типу ПЗ даний проект належить до системи автоматизованих розрахунків, а значить його трудомісткість складає 414 людино/г.

Розрахунок трудомісткості розробки технічного проекту розраховується за наступній формулою 5.2.

$$T_{ТЗ} = T_y * L_1 * K_n, \quad (4.2)$$

де T_y – укрупнена форма часу на розробку аналога ПЗ, людино/г, яка корегується поправочним коефіцієнтом K_n , враховуючи умови розробки за допомогою комп'ютеру ($K_n = 0.7$).

L_1 – питома вага даного етапу розробки з урахуванням ступеню новизни становить 0.1.

K_n – поправочний коефіцієнт, враховуючи ступінь новизни ($K_n = 0.7$).

$$T_{ТЗ} = (414 * 0.7) * 0.1 * 0.7 = 20.3 \quad (\text{людина} - \text{дні}).$$

Розрахунок трудомісткості розробки технічного проекту проводиться за формулою 5.3.

$$T_{ТП} = T_y * L_2 * K_n, \quad (4.3)$$

де L_2 – питома вага даного етапу розробки з урахуванням ступеню новизни становить 0.15.

$$T_{ТП} = (414 * 0.7) * 0.15 * 0.7 = 30,4 \quad (\text{людина} - \text{дні}).$$

					KPM.KI.0.884-03.2.3	Арк.
						58
Змн.	Арк.	№ докум.	Підпис	Дат		

Розрахунок трудомісткості розробки робочого проекту проводиться за формулою 5.4.

$$T_{PI} = T_y * L_3 * \kappa_n * \kappa_T, \quad (4.4)$$

де L_3 – питома вага даного етапу розробки з урахуванням ступеню новизни становить 0.7.

κ_T – поправочний коефіцієнт, враховуючи ступінь використання в розробці типових програм ($\kappa_T = 0.8$).

$$T_{PI} = (414 * 0.7) * 0.55 * 0.7 * 0.8 = 89,3 (\text{людина} - \text{дні}).$$

Розрахунок трудомісткості впровадження проводиться за формулою 4.5

$$T_{BH} = T_y * L_4 * \kappa_n, \quad (4.5)$$

де L_4 – питома вага даного етапу розробки з урахуванням ступеню новизни становить 0.2.

$$T_{BH} = (414 * 0.7) * 0.2 * 0.7 = 40,6 \quad (\text{людина} - \text{дні}).$$

В таблиці 5.3 приведено розрахунок трудомісткості розробки проекту
Тривалість розробки ПП визначається за формулою 5.6.

$$T_{III} = \sum T_{ij} / (5 * 0.73 * 360), \quad (4.6)$$

- де $\sum T_{ij}$ – сумарна тривалість розробки, г;
- 6 – тривалість робочого дня, г;
 - 0.73 – коефіцієнт переведу у календарні дні;
 - T_{ij} – тривалість виконання j-го виду роботи по i-му етапу.

Таблиця 4.3 – Розрахунок трудомісткості проекту

Назва етапу	Розрахунок, дні
1.Технічне завдання	$T_{T3} = 20,3$, $T_{KK} = 0.7 * N_{T3} = 0.7 * 20,3 = 14,21$, $T_{HK} = 0.15 * N_{T3} = 0.15 * 20,3 = 3.05$
2.Розробка технічного проекту	$T_{TI} = 30,4$, $T_{KK} = 0.7 * N_{TI} = 0.7 * 30,4 = 21,28$, $T_{HK} = 0.15 * N_{TI} = 0.15 * 30,4 = 4,56$
3.Розробка робочого	$T_{PI} = 89,3$, $T_{KK} = 0.7 * N_{PI} = 0.7 * 89,3 = 62,51$,

проекту	$T_{HK} = 0.15 * N_{PI} = 0.15 * 89,3 = 13.395$
4.Відладка та впровадження	$T_{BH} = 40,6, T_{KK} = 0.7 * N_{BH} = 0.7 * 40,6 = 28,42,$ $T_{HK} = 0.15 * N_{BH} = 0.15 * 40,6 = 6,09$
Усього:	$\sum T_{ij} = 20,3 + 14,21 + 3,05 + 30,4 + 21,28 + 4,56 + 89,3 + 62,51 + 13.395 + 40,6 + 28,42 + 6,09 = 334,115$

$$T_{III} = 334,115 / (5 * 0.73 * 360) = 334,115 / 1314 = 0,2542(\text{років}).$$

$$T_{III} = 0,2542 * 360 = 91,512(\text{днів})$$

4.3.2 Визначення ціни проекту

Для визначення ціни необхідно розрахувати основну заробітну плату працівників, матеріальні витрати, вартість машино-години та інші витрати на розробку продукту. Розрахунок проводиться за формулою 5.7.

$$Ц = K * C + \Pi_p, \quad (4.7)$$

– K – коефіцієнт обліку витрат на виготовлення дослідного зразка ПП як продукції виробничого технічного призначення ($K = 1.1$).

– C – витрати на розробку продукції (кошторисна собівартість).

– Π_p – нормативний прибуток.

Розрахунок нормативного прибутку проводиться за формулою 5.8.

$$\Pi_p = (C - C_M) * P_n / 100, \quad (4.8)$$

– C_M – матеріальні витрати, грн.

– P_n – норматив рентабельності приймається рівним 25%.

$$\Pi_p = (47474 - 3550) * 0,25 = 1098,1$$

Розрахунок матеріальних витрат приведено в таблицях 5.4

Таблиця 4.4 – Розрахунок витрат на відрядження

Найменування матеріалу	Кількість	Ціна за одиницю, грн	Сума, грн
Канцтовари			550
Папір	15	60	900
Виготовлення ксерокопій креслень			950

					KPM.KI.0.884-03.2.3	Арк.
						60
Змн.	Арк.	№ докум.	Підпис	Дат		

Виготовлення рекламних листівок	100	2	200
Дискети, CD- диски, Flash- накопичувачі			950
Всього			3550

$$Ц = 1.1 \cdot 47474 + 1098,1 = 53319,5$$

Основна заробітна плата враховує основну заробітну плату виконавця, безпосередньо зайнятого розробкою даного ПП, з урахуванням його посадового окладу та часу участі в розробці. Розрахунок виконується за формулою 4.9.

$$C_{zo} = \sum (Z_i * K_0 / Dp) * \tau_i, \quad (4.9)$$

Де:

- Z_i – середньомісячний оклад і-того виконавця, грн.
- K_0 – коефіцієнт обліку окладу керівників і консультантів проекту приймається рівним 0.1.
- Dp – середня кількість робочих днів в місяці приймається рівною 21 дню.
- τ_i – трудомісткість робіт, що виконуються і-тим виконавцем, людино/дні.

У розробці задіяні постановник задачі та розробник проекту (мережі), середньомісячний оклад яких складає 10800, 5000 грн. відповідно.

Трудомісткості робіт складають 91,512 чол-дні.

$$C_{zo} = \frac{10800 * 0.1 * 40}{21} + \frac{5000 * 91,512}{21} = 2057,1 + 21788,6 = 23845,7.$$

Розрахунок додаткової заробітної платні враховує всі виплати безпосереднім виконавцям за час непропрацьований на виробництві, у тому числі: оплата чергових відпусток, компенсації за недовикористану відпустку, оплати пільгового годинника підліткам та ін. і проводиться за формулою (4.10).

$$C_{zd} = C_{zo} * K_d, \quad (4.10)$$

					KPM.KI.0.884-03.2.3	Арк.
						61
Змн.	Арк.	№ докум.	Підпис	Дат		

де K_d – коефіцієнт відрахувань на заробітну платню приймається рівним 0.1.

$$C_{zd} = 23845,7 * 0.1 = 2385,57(\text{грн}).$$

Розрахунок відрахувань на соціальне страхування враховує відрахування до бюджету соціального страхування за встановленим державою тарифом від суми основної і додаткової заробітної платні і розраховується за формулою (5.11).

$$C_{cc} = K_{cc} * (C_{zo} + C_{zd}), \quad (4.11)$$

Де: K_{cc} – коефіцієнт відрахувань на соціальне страхування приймається рівним 0,22, він є єдиним соціальним несклом.

$$C_{cc} = 0,22 * (23845,7 + 2385,57) = 5770,87(\text{грн})$$

Загальновиробничі витрати розраховуються за формулою 4.12 і враховують витрати на загальногосподарські витрати, позавиробничі витрати і витрати на управління.

$$C_H = K_H * C_{zo}, \quad (4.12)$$

де K_H – коефіцієнт накладних витрат приймається рівним 0.5.

$$C_H = 0.5 * 23845,7 = 11922,85(\text{грн}).$$

В таблиці 4.5 представлена кошторисна собівартість ПП.

Таблиця 4.5 – Сметна вартість ПП

Назва категорії витрат	Кошторисна собівартість, грн.
Основна заробітна плата	23845,7
Додаткова заробітна плата	2384,57
Відрахування на соціальне страхування	5770,87
Матеріальні витрати	3550

Загальновиробничі витрати	11922,85
Сумарна собівартість	47474

4.3.3 Визначення капітальних і поточних витрат

Розрахунок капітальних витрат, пов'язаних з впровадженням (вдосконаленням) ІС здійснюється за формулою 5.13.

$$K_2 = K_{\Pi} + K_{KO} + K_{до} + K_B, \quad (4.13)$$

- K_{Π} – передвиробничі витрати;
- K_{KO} – вартість комп'ютерного устаткування(
- $K_{до}$ – вартість допоміжного устаткування, необхідного для надійної роботи мережі доступу;
- K_B – вартість будівництва (реконструкції) у зв'язку з впровадженням мережі доступу.

Передвиробничі витрати – всі витрати, пов'язані з проектуванням, розробкою, відладкою та впровадженням програмного забезпечення, навчання обслуговуючого систему персоналу, перепідготовка частини персоналу підприємства та інші передвиробничі витрати розраховуються за формулою 4.14.

$$K_{\Pi} = Ц * 1.0. \quad (4.14)$$

Таким чином, K_{Π} приймаються у розмірі 100% від вартості розробленого ПП і становлять:

$$K_{\Pi} = 53319,5 * 1.0 = 53319,5(\text{грн}).$$

Спеціальне комп'ютерного устаткування вартістю 121000 грн, додаткове 8900 грн. Будівництво, пов'язане з впровадженням проекту приблизно 15000 грн. З цього виходить, що величина капітальних витрат складає:

$$K_2 = 53319,5 + 121000 + 8900 + 15000 = 198219,5(\text{грн}).$$

Розрахунок поточних витрат, пов'язаних з впровадженням інформаційної системи здійснюється за формулою 5.15. Основними користувачами програми є

					KPM.KI.0.884-03.2.3	Арк.
						63
Змн.	Арк.	№ докум.	Підпис	Дат		

адміністратор та інженер самого проекту, в разі модернізації. Тобто при модернізації скорочено одного робітника.

$$C = C_{\text{опл}} + C_A + C_{\text{ел}} + C_P + C_{\text{доп}} + C_{\text{п}}, \quad (4.15)$$

Де:

- $C_{\text{опл}}$ – річний фонд основної і додаткової оплати праці персоналу, обслуговуючого мережі;
- C_A – сума річних амортизаційних відрахувань від вартості основного і допоміжного устаткування мережі;
- $C_{\text{ел}}$ – вартість витрат на електроенергію за рік;
- C_P – вартість річного ремонту основного і допоміжного устаткування;
- $C_{\text{доп}}$ – річна вартість допоміжних матеріалів, пов'язаних з експлуатацією мережі;
- $C_{\text{п}}$ – вартість річного утримання приміщень.

Модернізована мережа дозволить скоротити робітника, залишиться 2 робітники, з оплатою 45 грн./год кожен. Тобто виходить економія за рік 82125 грн на одному робітнику, який був скорочений.

Річний фонд основної заробітної платні персоналу, обслуговуючого мережу розраховується за формулою 5.16.

$$Z_{\text{опл}} = \sum \chi_{ci} * Z_{ci} + \sum \chi_{pj} * t_{cj} * \Phi_{pj}, \quad (4.16)$$

- χ_{ci}, χ_{pj} – чисельність, відповідно, фахівців і-ї категорії та j-того розряду, обслуговуючих ІС;
- t_{cj} – годинна тарифна ставка робочого j-того розряду;
- Φ_{pj} – річний фонд робочого часу j-того розряду приймається рівним 1825 годин.

До модернізації:

$$Z_{\text{опл}} = 3 \cdot 45 \cdot 1825 = 246375 (\text{грн.})$$

Після модернізації:

					KPM.KI.0.884-03.2.3	Арк.
						64
Змн.	Арк.	№ докум.	Підпис	Дат		

$$З_{опл} = 2 \cdot 45 \cdot 1825 = 164250(\text{грн.})$$

Фонд додаткової заробітної платні розраховується за формулою 4.17.

$$З_{доп} = З_{осн} * K_{доп}, \quad (4.17)$$

де $K_{доп}$ – коефіцієнт додаткової заробітної платні, приймається рівним 0.1.

До модернізації:

$$З_{доп} = 246375 * 0.1 = 24637,5(\text{грн.})$$

Після модернізації:

$$З_{доп} = 164250 * 0.1 = 16425(\text{грн.}).$$

Розрахунок нарахувань на заробітну платню проводиться за формулою 4.18.

$$З_{нар} = (З_{осн} + З_{доп}) * K_B, \quad (4.18)$$

де K_B – коефіцієнт відрахувань на соціальні потреби, приймається рівним 0.22.

До модернізації:

$$З_{нар} = (246375 + 24637,5) * 0.22 = 59842,75(\text{грн.})$$

Після модернізації:

$$З_{нар} = (164250 + 16425) * 0.22 = 39708,9(\text{грн.}).$$

Таким чином, загальні витрати на оплату праці розраховуються за формулою 5.19.

$$C_{опл} = З_{осн} + З_{доп} + З_{нар}, \quad (4.19)$$

До модернізації:

$$C_{опл} = 246375 + 24637,5 + 59842,75 = 330855,25(\text{грн.})$$

Після модернізації

$$C_{опл} = 164250 + 16425 + 39708,9 = 220383,9(\text{грн.}).$$

Розрахунок амортизаційних відрахувань розраховується за формулою 4.20.

$$C_a = K_{ко} * H_a / 100, \quad (4.20)$$

Де H_a – норма амортизаційних відрахувань, приймається рівної 60% для комп'ютерного устаткування.

$$C_a = 121000 * 60 / 100 = 72600(\text{грн.}).$$

					KPM.KI.0.884-03.2.3	Арк.
						65
Змн.	Арк.	№ докум.	Підпис	Дат		

Річна вартість споживаної електроенергії визначається за формулою 4.21.

$$C_{EL} = M_y * T_{KO} * C_o * K_n, \quad (4.21)$$

- M_y – встановлена сумарна потужність комп'ютерного устаткування, приймається рівною 5.2 кВт, хоча раніше була 8.1 кВт;
- T_{KO} – річний фонд роботи ЕОМ з урахуванням часу на профілактичні огляди складає $220*6=1320$ годин, хча раніше було $245*6=1470$;
- C_o – вартість 1 кВт електроенергії на даний момент складає 3,05грн.;
- K_n – коефіцієнт інтенсивного використання потужності, приймається рівним 0.9.

Після модернізації

$$C_{EL} = 8,1*1470*3,05*0.9 = 32684,72(\text{грн})$$

Після модернізації

$$C_{EL} = 5,2*1320*3,05*0.9 = 18841,68(\text{грн}).$$

Вартість річного ремонту основного і допоміжного устаткування становить 6 % від вартість комп'ютерного і допоміжного устаткування і складає:

$$C_P = K_{KO} * 0.06 = 121000 * 0.06 = 7260(\text{грн}).$$

Річна вартість допоміжних матеріалів, пов'язаних з експлуатацією ІС становить 1.5 % від вартості комп'ютерного і допоміжного устаткування і складає:

$$C_{BSP} = K_{KO} * 0.015 = 121000 * 0.015 = 181,5(\text{грн}).$$

Витрати на виробниче приміщення складають 1950 грн. в рік за 1 м², раніше було 19 м²:

$$C_{II} = 19*1950 = 37050(\text{грн})$$

а зараз 12 м² и складає:

$$C_{II} = 12*1950 = 23400(\text{грн}).$$

До впровадження проекту витрати складали:

					KPM.KI.0.884-03.2.3	Арк.
						66
Змн.	Арк.	№ докум.	Підпис	Дат		

$$C_1 = 330855,25 + 72600 + 32684 + 7260 + 181,5 + 37050 = 480630,75(\text{грн})$$

Таким чином, поточні витрати після впровадження даного проекту складають:

$$C_2 = 220383,9 + 72600 + 18841,68 + 7260 + 181,5 + 23400 = 321607,08(\text{грн})$$

4.3.4 Визначення показників економічної ефективності проекту

Очікуваний економічний ефект розраховується за формулою 4.22.

$$E_o = E_z - E_n * K_n, \quad (4.22)$$

- E_r – річна економія на поточних витратах;
- E_n – нормативний коефіцієнт ефективності єдиноразових витрат – 0.25;
- K_n – єдиноразові витрати на проект.

Річна економія складається з поточних витрат та приросту прибутку у зв'язку із впровадженням проекту, та обчислюється за формулою 4.23.

$$E_z = (C_1 - C_2) + \Delta\P, \quad (4.23)$$

– C_1, C_2 – відповідно поточні витрати, відповідно до та після впровадження проекту (грн);

– $\Delta\P$ – приріст прибутку господарюючого суб'єкта або його структурного підрозділу при впровадженні проекту (грн) в даному випадку неможливо підрахувати приріст прибутку, тому він не буде рахуватися.

$$E_z = (480630,75 - 321607,08) = 159023,67(\text{грн})$$

Тобто було зекономлено 159023,67 грн.

$$E_o = 159023,67 - 0,25 * 321607,08 = 80401,77$$

Розраховуємо коефіцієнт ефективності єдиноразових витрат за формулою 4.24

$$E = E_z / K_n, \quad (4.24)$$

$$E = E_z / K_n = 159023,67 / 198219,5 = 0,802$$

$E > E_n$, тому проект економічно ефективний.

Термін окупності одноразових витрат обчислюється за формулою 4.25

$$T = 1 / E, \quad (4.25)$$

					KPM.KI.0.884-03.2.3	Арк.
						67
Змн.	Арк.	№ докум.	Підпис	Дат		

$$T = 1/0,802 = 1,25 \vec{\text{ (років)}}$$

$$T = 1,25 \cdot 365 = 465 \text{ (дні)}$$

Техніко-економічні показники проекту, , відображені в таблиці 4.6

Таблиця 4.6 – Техніко-економічні показники проекту

/п	Найменування показника	Одиниця виміру ювання	Значення показника
			після впровадження проекту
	Час розробки проекту	дні	91,512
	Ціна ПП	грн.	53319,5
	Капітальні витрати	грн.	198219,5
	Поточні витрати	грн/рік	321607,08
	Економічний ефект від реалізації проекту	грн/рік	159023,67
	Термін окупності	Рік	1,25

Висновки:

Усі отримані показники знаходяться в рамках норми. Впровадження проекту є економічно ефективним. Після впровадження проекту значно зменшились час виконання даної роботи, а також підвищилась ефективність за рахунок оновлення обладнання та модернізації мережі взагалі. Термін окупності складає приблизно 465 дні. Що також каже про великі можливості проекту та можливості його подальшого розвитку. Завдяки модернізації за рік експлуатації мережі буде зекономлено 159023 грн.

					KPM.KI.0.884-03.2.3	Арк.
						68
Змн.	Арк.	№ докум.	Підпис	Дат		

РОЗДІЛ 5

ОХОРОНА ПРАЦІ ТА БЕЗПЕКА ЖИТТЄДІЯЛЬНОСТІ, ТЕХНІКА БЕЗПЕКИ, ЕКОЛОГІЯ

Тема даного дипломної роботи є «Модернізація мережі доступу у селищі Холодна Балка із аналізом топологічної структури». У зв'язку із цим у розділі охорони праці будемо розглядати офісне приміщення, в якому буде проводитись аналіз, розрахунки та моделювання, з наступними характеристиками: 12х9х3,5м.

$$12 \cdot 9 = 108(m^2)$$

$$12 \cdot 9 \cdot 3,5 = 378(m^3)$$

Виходить: 108 м²; 378 м³. За нормативними даними на одну людину має бути не менш 20 м³ та не менш 6 м². Тому максимальна кількість людей у цьому приміщенні 18.

5.1 Шкідливі та небезпечні фактори в роботі у робочій зоні

До небезпечних і шкідливих виробничих факторів у зоні робочих місць експлуатуючі системи з комп'ютерами відносять: статична електрика, електромагнітні випромінювання, шум, вібрації, недостатнє висвітлення, вентиляція та ін. Персонал піддається також і впливу шкідливих психофізіологічних виробничих факторів, які у свою чергу підрозділяються по характері дії на фізичні перевантаження й нервово-психічні перевантаження. Фізична й нервово-психічні: розумова перенапруга; монотонність праці; емоційні перевантаження.

Професійні шкідливості – несприятливі для здоров'я фактори трудового (виробничого) процесу або незадовільні санітарно-гігієнічні умови.

					KPM.KI.0.884-03.2.3	Арк.
						69
Змн.	Арк.	№ докум.	Підпис	Дат		

Професійно шкідливий чинник – виробничий фактор, дія якого в певних умовах може мати несприятливий вплив на працездатність і стан здоров'я працівника аж до виникнення професійного захворювання.

Професійне захворювання – хвороба, викликана дією специфічного шкідливого фактора в умовах виробництва, яка підтверджена в установленому порядку.

Специфічний чинник виробничого середовища – фактор виробничого середовища, який не може бути виключений з виробничого середовища без зупинки процесу виробництва.

Залежно від часу і інтенсивності впливу на працівника, виробничі фактори можуть бути небезпечними або шкідливими.

При миттєвій дії фактор стає небезпечним, а при тривалому впливі — шкідливим.

Небезпечним називають виробничий фактор, вплив якого на організм працюючого у відповідних умовах праці може призвести до травм або іншого раптового, різкого погіршення стану здоров'я.

Шкідливим називають виробничий фактор, вплив якого на організм працюючого може призводити в певних умовах до захворювання або зниження рівня працездатності.

Однією із основних цілей охорони праці на підприємстві є оцінка обстановки та характеристик трудового процесу в частині його впливу на здоров'я і життя працівника.

Для досягнення цього завдання державою встановлено низку критеріїв оцінки, які допомагають визначити ступінь небезпечності умов праці на підприємствах, що використовують працю найманих робітників.

Класифікація виробництва по ступені пожежної, вибухової і вибухопожежної безпеки

По ступені вибуховий, вибухопожежної і пожежної безпеки приміщення із ВДТ ЕОМ ставляться до категорії Д, у яких використовуються негорючі речовини й матеріали в холодному стані.

					КРМ.КІ.0.884-03.2.3	Арк.
						70
Змн.	Арк.	№ докум.	Підпис	Дат		

5.2 Методи зниження шкідливих та небезпечних факторів в роботі

Вимоги електробезпеки при експлуатації ВТД ЕОМ

Лінія електромережі для ЕОМ виконується як окрема групова трьохпровідна мережа, шляхом прокладки фазового, нульового робочого, нульового захисного проводів.

Нульовий захисний провід використовується тільки для занулення електроприймача.

Використання нульового робочого проводу в якості нульового захисного забороняється.

Штепсельні з'єднання й розетки повинні мати спеціальні контакти для підключення нульового захисного проводу.

Не допускається підключення ПЕОМ до звичайного двохпровідної мережі, у тому числі з використанням перехідних пристосувань. Є неприпустимим експлуатація кабелів і проводів з ушкодженою ізоляцією, саморобних подовжувачів.

Захисне заземлення - навмисне з'єднання металевих неструмоведучих частин, які можуть виявитися під напругою зі штучним заземленням. Застосовується в мережах з напругою до 1000 В з ізольованої нейтраллю, і вище 1000 У з будь-яким режимом нейтралі.

За способом захисту людини від поразки електричним струмом ЕОМ повинні відповідати першому класу захисту.

Розрахунок заземлюючого пристрою

Вихідні данні:

- d - зовнішній діаметр труби $d = 0.04\text{м}$;
- Довжина вертикальних заземлювачей $l = 2,2\text{ м}$; Відношення $l/l' = 2$;
- ρ_{ϕ} – питомий опір ґрунту суглинки в місці пристрою заземлення $\rho_{\phi} = 100\text{ Ом}\cdot\text{м}$;

					КРМ.КІ.0.884-03.2.3	Арк.
						71
Змн.	Арк.	№ докум.	Підпис	Дат		

– ϕ - безрозмірний кліматичний коефіцієнт, що враховує сезонні коливання вологості ґрунту 1,3.

Розрахункове значення питомого опору ґрунту обчислюємо по формулі:

$$\rho_p = \rho_\phi \cdot \phi \quad (5.1)$$

Де:

- ρ_ϕ – питомий опір ґрунту в місці пристрою заземлення ;
- ϕ - безрозмірний кліматичний коефіцієнт, що враховує сезонні коливання вологості ґрунту

$$\rho_p = 100 \cdot 1,3 = 130 \text{ Ом} \cdot \text{м}$$

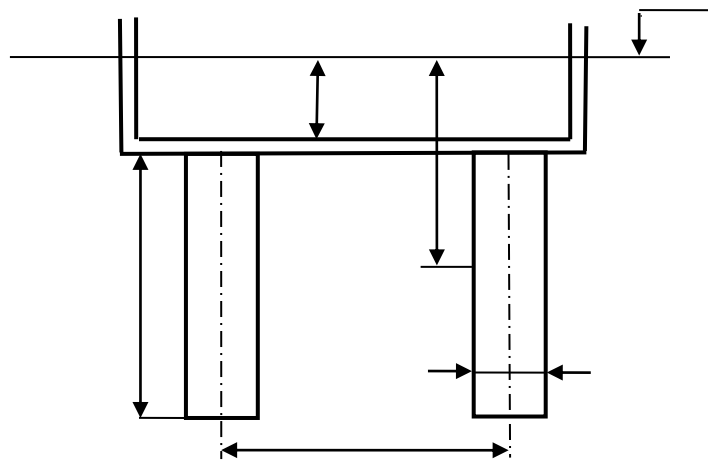


Рис. 5.1 – Устройство заземлителей

$$t_0 = 0,5 \text{ м}, l = 2,2 \text{ м}$$

l - довжина заземлювачей, м;

l' - відстань між заземлювачами, м

$$\text{Тоді } t = t_0 + l/2 = 0,5 + 2,2/2 = 1,6 \text{ м}$$

Опір одного вертикального заземлювача обчислюємо по формулі:

$$R_{\text{тр}} = \frac{\rho_{\text{расч}}}{2\pi \cdot l} \left(\ln \frac{2 \cdot l}{d} + 0,5 \ln \frac{4t + l}{4t - l} \right), \quad (5.2)$$

$$R_0 = \frac{130}{2 \cdot 3,14 \cdot 2,2} \cdot \left(\ln \frac{2 \cdot 2,2}{0,04} + 0,5 \cdot \ln \frac{4 \cdot 1,6 + 2,2}{4 \cdot 1,6 - 2,2} \right) = 47,6 \text{ Ом}$$

					KPM.KI.0.884-03.2.3	Арк.
						72
Змн.	Арк.	№ докум.	Підпис	Дат		

Визначаємо необхідна кількість вертикальних заземлювачей:

$$n = \frac{R_0}{R_{mp}} = \frac{47,6}{4} = 11,9 \quad (5.3)$$

Де: $R_{тр} = 4 \text{ Ом}$ – необхідний опір.

Округляємо n до найближчого стандартного числа $n' = 20$ шт. Визначаємо загальний опір системи вертикальних заземлювачей:

$$R = \frac{R_0}{n' \cdot \eta_v} \quad (5.4)$$

Де: η_v – коефіцієнт використання вертикальних заземлювачей, тому що заземлювачі розташовані в ряд і відношення відстані між заземлювачами до їхньої довжини дорівнює 2, то за табличним значенням приймаємо $\eta_v = 0,67$ (Юдин «Охрана труда в машиностроении» с.270 таб.19)

$$R_{об} = \frac{47,6}{20 \cdot 0,67} = 3,55 \text{ Ом}$$

Визначаємо опір сполучної смуги. При розміщенні в ряд довжина смуги

$$L = l' \cdot (n' - 1) = 4,4 \cdot (20 - 1) = 83,6 \text{ м} \quad (5.5)$$

$$R_{пол} = \frac{\rho_p}{2\pi \cdot L \cdot \eta_r} \cdot \ln \frac{L^2}{d \cdot t_o}, \quad (5.6)$$

Де: η_r – коефіцієнт використання горизонтальних заземлювачей, приймаємо $\eta_r = 0,56$ (Юдин «Охороона праці в машинобудуванні» с.270 таб.20);

$$R_{пол} = \frac{100}{2 \cdot 3,14 \cdot 83,6 \cdot 0,56} \cdot \ln \frac{83,6^2}{0,04 \cdot 0,5} = 4,34 \text{ Ом}$$

Загальний опір системи обчислюється по формулі:

$$R = \frac{3,55 \cdot 4,34}{3,55 + 4,34} = 1,95 \text{ Ом}$$

$$R_c = \frac{R_{об} \cdot R_{пол}}{R_{об} + R_{пол}} (\text{Ом}) \quad (5.7)$$

Прийнята система заземлення задовольняє технічним вимогам, тому що загальний опір системи заземлення менше припустимого значення $R_{тр} \leq 4 \text{ Ом}$

					KPM.KI.0.884-03.2.3	Арк.
						73
Змн.	Арк.	№ докум.	Підпис	Дат		

Пожежна профілактика

Пожежі становлять більшу небезпеку для працюючих і можуть заподіяти величезний матеріальний збиток. Пожежна безпека може бути забезпечена мірами пожежної профілактики й активного пожежного захисту. Поняття пожежної профілактики включає комплекс заходів, необхідних для попередження виникнення пожежі або зменшення його наслідків. Під активним пожежним захистом розуміються міри, що забезпечують успішну боротьбу з виникаючими пожежами або вибухонебезпечною ситуацією.

Причини пожеж в електроустановках

Пожежі в електроустановках відбуваються через:

- короткого замикання;
- перевантаження мереж;
- більших перехідних опорів;
- від електронагрівальних приладів.

Перевантаження мереж відбувається в результаті проходження по них напруги, що перевищує номінальний. Таке може відбутися у випадку підключення великої кількості споживачів. Це приведе до руйнування, плавлення й згоряння ізоляції, що спричиняє коротке замикання.

Коротке замикання відбувається в тому випадку, коли крапки різних фаз мережі з'єднуються через малий опір. Внаслідок чого миттєво збільшується струм, відбувається виділення великої кількості тепла.

Міри захисту:

- дотримання нормальних режимів експлуатації;
- своєчасне проведення регламентних робіт;
- застосування плавких запобіжників і автоматів.

Більшу роль у пожежонебезпеці грає правильний вибір використання електроустаткування.

					КРМ.КІ.0.884-03.2.3	Арк.
						74
Змн.	Арк.	№ докум.	Підпис	Дат		

Система висвітлення з лампами накаливання найнебезпечніші через те, що температура поверхні колби лампи приблизно дорівнює 500 °С. Більшу роль у безпеці грає правильний вибір типу світильника.

Засоби пожежогасіння

Як засоби пожежогасіння на даному об'єкті застосовуються вуглекислотні вогнегасники, призначені для гасіння загорянь установок напругою до 1000 В: ручні ВВ-2А, ВВ-5, ВВ-8.

У вуглекислотних вогнегасниках застосовують зріджений двоокис вуглецю. Вогнегасяща дія його полягає в розведенні повітря й зниженні в ньому змісту кисню до концентрації, при якій припиняється горіння. Вогнегасящий ефект обумовлюється втратами теплоти на нагрівання двоокису вуглецю й зниженням теплового ефекту реакції.

Розрахунок вуглекислотної установки для гасіння пожежі в приміщенні

Кількість вогнегасящого газового складу G_r (кг) визначається за залежністю

$$G_r = 1,25 \cdot (G_v \cdot V_n \cdot K_y), \text{ кг} \quad (5.8)$$

Де:

- G_v - вогнегасяща концентрація газового складу для вуглекислоти;
- ($G_v = 0,7 \text{ кг/м}^3$);
- V_n – обсяг приміщення, що захищається, м^3 ;
- K_y – коефіцієнт, що враховує особливості процесу газообміну, джерела вуглекислоти крізь нещільності й щілині приміщення, що захищається, приймається 1...1,2

$$G_r = 1,25 \cdot (0,7 \cdot 379 \cdot 1) = 331,625 \text{ кг}$$

Потрібна кількість робочих балонів N_6 (од.) з вуглекислотою визначається як

$$N_6 = \frac{G_2}{V_6 \cdot \rho \cdot \alpha_n}, \text{ од.} \quad (5.9)$$

Де:

- $V_6 = 40 \text{ л}$ – обсяг балону;

					КРМ.КІ.0.884-03.2.3	Арк.
						75
Змн.	Арк.	№ докум.	Підпис	Дат		

- $\rho = 0,625$ кг/л – щільність засобу гасіння;
- $\alpha_v = 1$ – коефіцієнт наповнення.

$$N_6 = \frac{331,625}{40 \cdot 0,625 \cdot 1} \approx 13, \text{ од.}$$

Кількість резервних балонів варто прийняти такий же як число робочих балонів. $N_6 = N_p = 13$ балонів. Загальна кількість балонів 26 шт.

Виробнича санітарія – це система організаційних та технічних заходів, які спрямовані на усунення потенційно небезпечних факторів і запобігання професійних захворювань та отруєнь.

До організаційних заходів належать:

- дотримання вимог охорони праці жінок та осіб віком до 18 років;
- проведення попередніх та періодичних медичних оглядів осіб, які працюють у шкідливих умовах;
- забезпечення працюючих у шкідливих умовах лікувально-профілактичним обслуговуванням тощо.

Технічні заходи передбачають:

- систематичне підтримання чистоти у приміщеннях і на робочих місцях;
- розробку та конструювання обладнання, що вилучає виділення пилу, газів та пари, інших шкідливих речовин у виробничих приміщеннях;
- забезпечення санітарно-гігієнічних вимог до повітря виробничого середовища;
- улаштування систем вентиляції та кондиціонування робочих місць зі шкідливими умовами праці;
- забезпечення захисту працюючих від шуму, ультра- та інфразвуку, вібрації, різних видів випромінювання.

Таким чином, запобігання професійних захворювань і отруєнь здійснюється через здійснення комплексу організаційних і технічних заходів, які спрямовані на оздоровлення повітряного середовища, виконання вимог гігієни та особистої безпеки працюючих.

Розрахунок вентиляції приміщення

					KPM.KI.0.884-03.2.3	Арк.
						76
Змн.	Арк.	№ докум.	Підпис	Дат		

Розрахувати продуктивність системи вентиляції в приміщенні із заданими параметрами, з урахуванням максимального розташування робочих місць із персональними комп'ютерами.

Визначаємо площу приміщення:

$$S = A \times B = 12 \times 9 = 108 \text{ м}^2 \quad (5.10)$$

З обліком, що кількість людей ($n_{\text{л}}$) і кількість ЕОМ ($n_{\text{к}}$) дорівнює 23, визначимо кількість тепла виділюване людьми:

$$Q_{\text{л}} = n_{\text{л}} \times q_{\text{л}}; \quad (5.11)$$

$q_{\text{л}}$ – питома тепловиділення 1 чоловік (170 Вт);

$$Q_{\text{л}} = 18 \times 170 = 3060 \text{ Вт};$$

Визначимо кількість тепла виділюване від устаткування:

$$Q_{\text{об}} = n_{\text{к}} \times q_{\text{к}}; \quad (5.12)$$

$q_{\text{к}}$ – питома тепловиділення 1 комп'ютера (400 Вт);

$$Q_{\text{об}} = 18 \times 400 = 7200 \text{ Вт};$$

Визначимо кількість тепла, що виділяє освітлення:

$$Q_{\text{осв}} = S_{\text{пр}} \cdot q_{\text{осв}}; \quad (5.13)$$

Де:

– $S_{\text{пр}}$ – площа приміщення,

– $q_{\text{осв}}$ – кількість тепла, що приходить на 1 м² площини приміщення від штучного освітлення.

$$Q_{\text{осв}} = 108 \times 10 = 1080 \text{ Вт};$$

Визначимо кількість тепла, що надійшло через конструкції, що обгороджують:

$$Q_{\text{осв}} = V_{\text{пр}} \cdot q_{\text{пр}} \quad (5.14)$$

Де:

– $V_{\text{пр}}$ – об'єм приміщення,

– $q_{\text{осв}}$ – кількість тепла, що приходить на 1 м³ об'єму приміщення від навколишнього середовища.

					КРМ.КІ.0.884-03.2.3	Арк.
						77
Змн.	Арк.	№ докум.	Підпис	Дат		

$$Q_{\text{огр.кон}} = 108 \cdot 3,5 \cdot 18 = 6804 \text{ Вт};$$

Уважаємо загальне:

$$Q = Q_{\text{л}} + Q_{\text{об}} + Q_{\text{осв}} + Q_{\text{огр.кін}};$$

$$Q = 3060 + 7200 + 1080 + 6804 = 18144 \text{ Вт}$$

Об'ємна витрата повітря в приміщенні:

$$L = \frac{Q}{C_p \cdot \rho (t_{\text{выт}} - t_{\text{пр}})}, \quad (\text{м}^3/\text{ч}); \quad (5.15)$$

- ρ - щільність повітря ($1,2 \text{ кг/м}^3$);
- c - питома теплоємність повітря;
- $t_{\text{выт}}$, $t_{\text{пр}}$ – температура витяжного й припливного повітря.

$$L = \frac{18,14}{1,0 \cdot 1,2 \cdot (24 - 16)} = 1,89 \frac{\text{м}^3}{\text{с}} = 6804 \frac{\text{м}^3}{\text{ч}}$$

Визначимо потужність, споживану вентилятором, по формулі

$$N = \frac{L \Delta P_1 k \cdot 10^{-6}}{\eta_s \eta_n \cdot 3,6}, \quad [\text{Вт}] \quad (5.16)$$

Де:

- L - об'ємна витрата повітря, $[\text{м}^3/\text{з}];$
- ΔP - втрата тиску у повітроводі, $(300) [\text{Па}];$
- η_s - КПД вентилятора $(0,6);$
- η_n - КПД приводу, при приєднанні колеса через муфту $(0,9);$
- k - коефіцієнт запасу $(1,2);$

$$N = \frac{1,2 \cdot 6804 \cdot 300 \cdot 10^{-6}}{3,6 \cdot 0,6 \cdot 0,9} = 1,26 \text{ кВт}.$$

Розрахунок освітлення приміщення

Завданням розрахунку штучного освітлення є визначення потрібної потужності електричної освітлювальної установки для створення у виробничому приміщенні з персональними комп'ютерами заданої освітленості.

Підбираємо систему висвітлення для приміщення:

					KPM.KI.0.884-03.2.3	Арк.
						78
Змн.	Арк.	№ докум.	Підпис	Дат		

Довжина приміщення, $A = 12$ м

Ширина приміщення, $B = 9$ м

Висота приміщення, $H_n = 3,5$ м

E_n - нормована освітленість ($E_n = 350$)

k - коефіцієнт запасу ($k = 1,6$)

Визначимо площу приміщення:

$$S = A \times B = 12 \times 9 = 108 \text{ м}^2$$

Обчислимо висоту підвісу світильника над робочою поверхнею по формулі

$$H_p = H_n - h_{p.п.}, \quad (5.17)$$

Де:

– $h_{p.п.}$ – висота робочої поверхні;

– $H_p = 3,5 - 1 = 2,5$ м.

Для досягнення рівномірної освітленості необхідно, щоб відношення відстаней між центрами світильників до висоти їхнього підвісу над робочою поверхнею рівнялася конкретному числу, характерному для типу обраного світильника.

У приміщеннях із ВДТ ЕОМ обрані світильники серії ЛД. Для світильників ЛПО відношення $L_k/H_p = 1,4$

Визначимо L_k – відношення між центрами світильників розраховуємо по формулі

– $L_k = H_p \cdot 1,4$

– $L_k = 2,5 \cdot 1,4 = 3,5$ м

Визначимо кількість світильників:

$$N = A \cdot B / L_k^2 = 108 / 3,5^2 = 8,82 \quad (5.18)$$

Для визначення коефіцієнта використання світлового потоку необхідно обчислити індекс приміщення по формулі

$$i = \frac{a \cdot b}{H_{пав} \cdot (a + b)} \quad (5.19)$$

					КРМ.КІ.0.884-03.2.3	Арк.
						79
Змн.	Арк.	№ докум.	Підпис	Дат		

$$i = \frac{12 \cdot 9}{2.5 \cdot (12 + 9)} = 2,05$$

При $i = 2$ коефіцієнт використання світлового потоку $\eta = 57$.

Світловий потік одного світильника визначається формулою

$$\Phi_C = \frac{E_H \cdot k \cdot S \cdot z \cdot 100}{N \cdot \eta}, \quad (5.20)$$

Де:

- E_H - нормована освітленість ($E_H = 350$);
- k - коефіцієнт запасу ($k = 1,6$);
- S - площа освітлюваного приміщення;
- z - коефіцієнт нерівномірності висвітлення
- (для люмінесцентних ламп $z = 1,1$);
- N – число світильників (округлімо до 12 шт).

$$\Phi_C = \frac{350 \cdot 1,6 \cdot 108 \cdot 1,1 \cdot 100}{12 \cdot 57} = 9726 \text{ Лм}$$

Приймаємо до установки лампи ЛБ40 у кількості 3-х штук зі світловим потоком $\Phi = 3120 \text{ лм}$ у кожному світильнику.

$$\Delta = \frac{9360 - 9726}{9360} \cdot 100\% = -3,91\%$$

На практиці допускається відхилення світлового потоку реального від розрахункового в межах від -10% до +20%

Наша погрішність становить -3,91%. Підрахована погрішність задовольняє умові.

Розрахуємо потужність освітлювальної системи по формулі

$$P_c = N \cdot n \cdot P_1, \quad (5.21)$$

Де:

- N - кількість світильників;
- n - мінімальна кількість світильників;
- P_1 – мінімальна потужність.

					КРМ.КІ.0.884-03.2.3	Арк.
						80
Змн.	Арк.	№ докум.	Підпис	Дат		

$$P_c = 12 \cdot 3 \cdot 40 = 1440 \text{ Вт}$$

Загальні положення гігієнічних вимог до організації роботи із ВДТ ЕОМ

Облаштованість робочих місць ВДТ ЕОМ повинне забезпечити :

- належні умови висвітлення й відсутність відблисків;
- оптимальні параметри мікроклімату;
- належні ергономічні характеристики: наявність шуму й вібрації, рентгенівські випромінювання, електромагнітні випромінювання, електростатичне поле, наявність пилу, озону й оксиду азоту.

Вимоги до виробничих приміщень для експлуатації ВДТ ЕОМ

- Розміщення робочих місць із ВДТ ЕОМ у підвальних і цокольних поверхах заборонено.
- Мінімальна площа на одне робоче місце 6 м, мінімальний обсяг 20.
- Приміщення повинні бути обладнані системами природного й штучного висвітлення.
- Природне висвітлення повинне здійснюватися через світлові прорізи, орієнтовані переважно на північ, північний схід.
- Коефіцієнт природної освітленості (КЕО) повинен бути не нижче 1.5%.
- Приміщення із ВДТ ЕОМ не повинні граничити із приміщеннями, у яких рівень шуму й вібрацій перевищує припустимі норми.
- Віконні прорізи повинні бути обладнані регульованими пристосуваннями (жалюзі, штори, козирки, маркізи).
- Поверхня підлоги повинна бути рівної, не слизькою, з антистатичним покриттям.

Гігієнічні вимоги до параметрів виробничого середовища приміщень

					КРМ.КІ.0.884-03.2.3	Арк.
						81
Змн.	Арк.	№ докум.	Підпис	Дат		

Таблиця 5.1 - Мікроклімат приміщення

Пора року	Категорія роботи	Температура повітря, °C	Відносна вологість, %	Швидкість повітря в робочій зоні, м/с
холодне	легка 1а	22 – 24	40 – 60	0.1
	легка 1б	21 – 23	40 – 60	0.1
тепле	легка 1а	23 – 25	40 – 60	0.1
	легка 1б	22 - 24	40 – 60	0.2

Легка 1а - робота, сидячи, що не потребує більших фізичних витрат, тепловиділення порядку 140 Вт.

Легка 1б - робота, що виконується сидячи або пов'язана з невеликими переміщеннями, не потребуючих більших фізичних витрат, тепловиділення порядку 175 Вт.

Таблиця 5. 2 - Рівні іонізації повітря для приміщень із ВДТ ЕОМ:

Рівні	Число іонів в 1 повітря	
	n +	n –
мін необхідні	400	600
оптимальні	1500 – 3000	3000 – 5000
макс припустимі	50000	50000

Гігієнічні вимоги до організації й устаткування робочих місць із ВДТ ЕОМ

Робоче місце необхідно розташовувати щодо світлових прорізів, щоб природне висвітлення падало з боку, переважно ліворуч. Відстань від тильної поверхні одного до екрана іншого повинне бути не менш 2.5 м. Стандартна висота стола 680 - 800 мм.

Захист відстанню полягає в тім, щоб оператор перебував на відстані не менш 600 - 700 мм від монітора з урахуванням розмірів знаків і символів. Клавіатуру необхідно розташовувати на поверхні стола на відстані від 100 до

300мм від краю поверхні. Поверхня клавіатури повинна бути матової з коефіцієнтом відбиття 0,4.

Вимоги до режимів роботи й відпочинку при роботі із ВДТ ЕОМ. При виконанні робіт протягом дня, які ставляться до різних видів трудової діяльності. За основну роботу з ЕОМ варто вважати таку, котра займає не менш 50% робочого часу. Для програміста - рекомендується 15 хв. перерва через щогодини роботи. Для оператора ЕОМ при 8-ми часовому робочому дні передбачена всередині зміни перерва - 15 хвилин через кожні 2 години роботи.

Долікарняна допомога потерпілому. Перша долікарняна допомога при нещасних випадках від електричного струму складається з 2-х етапів: звільнення потерпілого від дії струму й надання йому медичної допомоги.

Міри першої медичної допомоги потерпілому від електричного струму залежать від його стану. У свідомості – але до цього був у непритомності або тривалий час перебував під струмом, йому необхідно забезпечити повний спокій до прибуття лікаря або терміново доставити в лікувальну установу.

При відсутності свідомості, але збереженому подиху й роботі серця потрібно рівно й зручно укласти потерпілого на м'яку підстилку, розстебнути пояс і одяг, забезпечити приплив свіжого повітря. Варто давати нюхати нашатирний спирт. Якщо потерпілий погано дихає - рідко, судорожно - або якщо подих поступово погіршується, у той час як при цьому триває нормальна робота серця, необхідно робити штучне дихання.

При відсутності ознак життя треба робити штучне дихання й непрямий масаж серця до появи ознак життя або до прибуття лікаря.

Висновок. Більшу частину свого життя людина проводить на роботі, де його можуть чекати не тільки приємні моменти, але й перевтома, стреси й виробничі травми. Тому дуже важливо створити необхідні умови для нормальної роботи людини й збереження його здоров'я під час трудової діяльності на підприємстві. Для цього й створена система охорони праці, що займається перевіркою дотримання всіх описаних вище вимог і розрахунком всіх технічних показників.

					КРМ.КІ.0.884-03.2.3	Арк.
						83
Змн.	Арк.	№ докум.	Підпис	Дат		

ВИСНОВКИ

Результатом дипломної роботи є дослідження процесу побудови мережі доступу, застосування методів розрахунку навантаження та автоматизація процесу розрахунку показників навантаження мереж доступу.

Основні завдання, вирішені в рамках роботи: Формування технічного завдання на проектування мережі доступу. Аналіз концепції мереж доступу, структура і функції мереж доступу. Проаналізовано методику розрахунку показників навантаження мереж доступу. Описані основні поняття та визначення. Проаналізовано процес розрахунку інтенсивності навантаження локального сегменту мережі доступу, розрахунок пропускної спроможності локального сегменту мережі доступу. Наведено приклади розрахунку навантаження в мережі доступу із використанням проаналізованої та обраної методики. Отримані результати сформовано в таблиці. Розрахунок сумарної інтенсивності питомого навантаження локального сегменту лінії доступу. Розрахунок пропускної спроможності лінії доступу локального сегменту. Розрахунок інтенсивності навантаження та пропускної спроможності транспортного сегменту мережі доступу. Оцінка пропускної спроможності вузлів доступу.

У розділі Реалізація автоматизованої системи розрахунку показників навантаження мереж доступу наведено Опис роботи програмного забезпечення; Опис інтерфейсу користувача; Опис системи для розробника; Технології реалізації автоматизованої системи.

Також виконано економічний розрахунок та розглянуто питання охорони праці.

					KPM.KI.0.884-03.2.3	Арк.
						84
Змн.	Арк.	№ докум.	Підпис	Дат		

СПИСОК ВИКОРИСТАНИХ ДЖЕРЕЛ

1. Konstan J. Riedl J. Sarwar B., Karypis G. Analysis of recommendation algorithms for e-commerce.ACM, pages 158-167, 2000.
2. Mohammed J. Zaki та Jr. Wagner Meira.Data Mining and Analysis: Fundamental Concepts and Algorithms. Cambridge University Press, May 2014 року.
3. Amedeo Napoli Laszlo Szathmary, Petko Valtchev. Generating rare asociation rules using the minimal rare itemsets family.J. Software and Informatics, 4 (3): 219-238, 2010.
4. Weiyang Lin, Sergio A. Alvarez, i Carolina Ruiz. Ефективна adaptive-support asociation rule mining для recommender systems.Data Mining and Knowledge Discovery, 6 (1): 83-105, 2002.
5. Chris Volinsky Yehuda Koren, Robert Bell. Matrix factorization techniques для recomender systems.IEEE COMPUTER, 42:30-37, 2009.
6. R. Agrawal, T. Imielinski, A. Swami. 1993. Mining Associations between Sets of Items in Massive Databases. In Proc. of the 1993 ACM-SIGMOD Int'l Conf. on Management of Data, 207-216.
7. R. Agrawal, R. Srikant. «Fast Discovery of Association Rules», In Proc. of the 20th International Conference on VLDB, Santiago, Chile, September 1994.
8. R. Srikant, R. Agrawal. «Mining Generalized Association Rules», In Proc. of the 21th International Conference on VLDB, Zurich, Switzerland, 1995.
9. R. Srikant, R. Agrawal. «Mining quantitative association rules in large relational tables». In Proceedings of the ACM SIGMOD Conference on Management of Data, Montreal, Canada, June 1996.
10. A. Savasere, E. Omiecinski, and S. Navathe, «An Efficient Algorithm for Mining Association Rules in Large Databases», In Proc. 21st Int'l Conf. Very Large Data Bases, Morgan Kaufmann, San Francisco, 1995.

					KPM.KI.0.884-03.2.3	Арк.
						85
Змн.	Арк.	№ докум.	Підпис	Дат		

11. J.S. Park, M.-S. Chen, and S.Y. Philip, «An Effective HashBased Algorithm for Mining Association Rules», In Proc. ACM SIGMOD Int'l Conf. Management of Data, ACM Press, New York, 1995.
12. S. Brin et al., «Dynamic Itemset Counting and Implication Rules for Market Basket Data», In Proc. ACM SIGMOD Int'l Conf. Management of Data, ACM Press, New York, 1997.
13. J. Hipp, U. Guntzer, and G. Nakaeizadeh. Algorithms for Association Rule Mining – A General Survey and Comparison. In Proc. ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, 2000.
14. Ricci F., Rokach L. Shapira B. Introduction to Recommender Systems Handbook, Springer, 2011, 35pp.
15. Resnick P., Varian H. R. «Recommender system». Communications of the ACM. 1997. № 3. 40. Pp. 56–58.
16. Linden G., Smith B.J. «Amazon.com recommendations: Item-to-Item collaborative filtering». IEEE Internet Computing. 2003. Vol. 7, Pp. 76–80.
17. Adomavicius B., Tuzhilin A. «Toward the Next Generation of Recommender Systems: A Survey of the State-of-the-Art and Possible Extensions». IEEE Transactions on knowledge and Data Engineering. 2005. Vol.17. Pp. 734-748.
18. Bobadilla J., Ortega A., Gutierrez H. A. «Recommender systems survey». Knowledge Based System. 2013.Vol. 46. Pp.109–132
19. Agrawal, R., Imielinski, T., Swami, A. N. «Mining association rules between sets of items in large databases» In Proceedings of the 1993 ACM SIGMOD International Conference on Management of Data. 1993. Pp. 207–216.
20. Agrawal R. Srikant R. «Fast algorithms for mining association rules» In Proc. 20th Int. Conf. Very Large Data Bases. 1994. Pp.487–499.
21. Han J., Pei J. 2000. Mining frequent patterns by pattern-growth: methodology and implications. ACM SIGKDD Explorations Newsletter . 2000. Vol. 2. Pp.14–20.
22. Zaki M.J. «Mining Non-Redundant Association Rules». Data Mining and Knowledge Discovery. 2004. Vol. 9. Pp. 223–248.

					<i>KPM.KI.0.884-03.2.3</i>	Арк.
						86
Змн.	Арк.	№ докум.	Підпис	Дат		