

Використання методів інтелектуального аналізу тексту для систем Recommender

Янелс Бетанкур¹Серхю Іларр²

¹Universidad de Zaragoza, Edificio Ada Byron, Mar'ia de Luna, 1, 50018 Caparoca, Іспанія
²IZA, Universidad de Zaragoza, Edificio Ada Byron, Mar'ia de Luna, 1, 50018 Caparoca, Іспанія

Ключові слова: системи рекомендацій, аналіз тексту, аналіз думок.

Анотація: Системи рекомендацій допомагають користувачам зменшити інформаційне перевантаження, в якому вони можуть страждати в поточну епоху Big Data, пропонуючи їм рекомендації в повільних елементах до їхнього смаку/вподобання та/або контекст (місце розташування, погода, час доби тощо). Ми стверджуємо, що методи інтелектуального аналізу тексту можна використовувати для розробки рекомендаційних систем. Таким чином, їх можна застосовувати для виявлення налаштувань користувача (профільованість користувача) а також для отримання контекстних даних. Для цього можна застосовувати текстовий майнінг на вмістах користувачів, текстових описах асоційованих елементів та інших текстах, написаних користувачем (наприклад, пости в соціальних мережах). У цій статті ми надаємо огляд робіт, що використовують методи інтелектуального аналізу тексту в області рекомендаційних систем, характеризуючи їх в повільних елементах до їх призначення та типу аналізованих текстових даних.

1 ВСТУП

В даний час рекомендаційні системи (RS) (Ricci et al., 2011 рік; Ricci et al., 2015) стали дуже популярними, оскільки вони можуть допомогти користувачам скоротити інформацію перевантаження, яке вони можуть зазнати в сучасну епоху великих даних, де потрібні передові методи керування даними. Ці системи надають рекомендації користувачам в повільних елементах до їхнього смаку/вподобання, дозволяючи їм фільтрувати велику кількість різних типів елементів (наприклад, музика, фільми, новини, книги, місця, послуги, торговельні площі, ресторани, точки кави тощо), так можуть мати для них особливе значення. Враховуючи їх поточну специфізацію, Системи рекомендацій продовжують залишатися гарячою областю для дослідження. У мобільних середовищах, зокрема, системи рекомендацій, які враховують не тільки уподобання користувача, а також їхній контекст (місцезнаходження, час, погода, умови руху тощо) мають особливе значення інтерес, який породив так звані контекстно-залежні рекомендаційні системи (CARS) (Adomavicius Тужиліс, 2011).

Ми стверджуємо, що використання технік інтелектуального аналізу тексту (Berry, 2004; Gupta et al., 2009) може бути використано для розробки систем рекомендацій. З різних точок зору. Популярні текстові завдання включають кластеризацію тексту, класифікацію тексту, вилучення інформації, резюмування тексту, аналіз настроїв/виявлення думок тощо. Наприклад, аналіз настроїв

Методи *usis* можуть бути застосовані для кількісного визначення уподобань користувачів на основі їхніх коментарів, висловлених природною мовою, або навіть для виявлення можливих невідповідностей (спам думок), якщо користувач надає як текстові коментарі, числовий рейтинг (видобуток думок про предмети). Крім того, їх можна використовувати для використання текстових описів різних елементів, рекомендувати (видобуток текстових даних елементів), наприклад, можливий аналіз інших текстів, написаних самим користувачем (наприклад, у блогах, соціальних мережах, його/її порядковий денний тощо) може розкривати інформацію про його/її уподобання, інтереси та навіть його/її контекст.

Релевантне доповнення до методів аналізу тексту полягає у використанні семантичних прийомів, як правило, заснованих на використанні онтологій. Як явна формальна специфікація спільної концептуалізації (Gruber et al., 1993), онтологія має стіть спільний словник, позначають види, властивості та взаємозв'язки цих поняття (Subramaniyaswamy та ін., 2019). Крім того, можливість використання онтологій як підтримки даних завдання видобутку, так як витяг інформації, онтології також можуть використовуватися безпосередньо рекомендаційною системою (Subramaniyaswamy та ін., 2019; Тарус та ін., 2018), наприклад, для встановлення зв'язку між користувачами та їхніми уподобаннями щодо теми рекомендацій (Tarus et al., 2018). На основі онтологій Системи рекомендацій, що використовують видобуток тексту, часто використовують а онтологію теми з пов'язаним контрольованим словником, визначати теми з текстових повідомлень, написаних

 <https://orcid.org/0000-0002-7073-219X>

уральською мовою (коментарі, описи, думки, рецензії тощо).

У цій статті ми подаємо огляд робіт, які використовують методи інтелектуального аналізу тексту в системі recommender (див. рис. 1). Структура решта цього документа виглядає наступним чином. У розділі 2 ми зосереджуємось на роботах у галузі рекомендаційних систем, які застосовувати техніку аналізу тексту на ввідних користувачів. У Розділі 3 ми розглядаємо системи рекомендацій, які виконують аналіз тексту на ввідних текстах, написаних користувачем. У розділі 4 ми розглядаємо роботи з використанням текстових даних про самі предмети. Нарешті, у розділі 5 ми підсумовуємо наш висновок та майбутню роботу.

2 РЕКОМЕНДАЦІЙНІ СИСТЕМИ

ВИКОРИСТАННЯ ТЕКСТОВОГО МАЙНІНГУ НА ВІДГУКИ КОРИСТУВАЧІВ

У цьому розділі ми зосередимося на аналізі тексту повторних переглядів користувачами (див. Таблицю 1). Іноді користувач висловлюють свої думки про продукти, які вони споживають, за допомогою текстових описів, розміщених у блогах, мейкблоггах (наприклад, веб-сайти з оглядами продуктів, чати, онлайн-системи зворотного зв'язку компанії, рейтингів користувачів або соціальних мереж (так як Facebook, Twitter, LinkedIn або Instagram)). Ці думки можуть стати дуже цінною інформацією для виявлення інтересів користувачів, які можуть бути використані для цілей рекомендацій (Рол Арора, 2019; Кім Чун, 2019).

Оскільки ці думки подано природною мовою, їх не можна використовувати безпосередньо для вивчення налаштування користувача та рекомендувати користувачеві інші елементи, що йому/їй, як очікується, сподобається. Тому типовий варіант спробувати використання видобутку тексту для систем рекомендацій для перетворення текстових ввідних користувачів у бали (наприклад, у діапазон від 0 до 5) або попередньо визначені категорії, наприклад позитивна думка, нейтральна думка або негативна думка, які можна використовувати для побудови матриць елементів користувача як системи рекомендацій використовують у процесі рекомендацій; зокрема, аналіз настроїв або методи аналізу думок (Liu and Zhang, 2012; Hegde Падма, 2017).

У (Aciar та ін., 2007) методи інтелектуального аналізу тексту є використовуються для перетворення думок у структуровані дані, використовуючи онтологію перекладу, яка охоплює як навички користувача та його досвід роботи з продуктом. Бали для кожної функції обчислюються для визначення рейтингу продуктів в дивізіоні до наявних коментарів про цей продукт. Система вимагає ввідного користувача в явного запиту рекомендації/оцінки щодо конкретного продукту, який їх цікавить, вони мають

щоб вибрати функції, які їх найбільше цікавлять. За межами При цьому слід зазначити, що цей тип рекомендацій не є персоналізованим, як індикатор дуальних переваг користувача не враховуються під час надання рекомендацій. Пропозиції враховують рівень досвіду рецензентів для забезпечення надійної рекомендації.

Надійшло в дугу в також є ключовим аспектом, який розглядається в інших пропозиціях. Таким чином, (Абуейн et al., 2017) представляє розробку та реалізацію системи рекомендацій на основі інформації, наданої довіреними людьми, використовуючи концепцію рівня довіри (LoT) (Abuein та ін., 2016). В (Рул Арора, 2019), також враховується довіра до рецензента, але в цьому випадку основна увага зосереджена на створенні короткого викладу кількох оглядів для певного продукту та достовірності (розраховується шляхом аналізу того, наскільки часто повторного перегляду виходять в довіряючих людях) використовуються для визначення важливості речення. Для автоматичного підсумовування ввідних користувачів, нечітко-тема кластеризація використовуються для групування схожих речень, а потім одне речення-представник з кожної групи вибирається для складання остаточного резюме.

Метою цієї роботи є створення репрезентативного резюме оглядів (Twitter), яке буде надано разом з рекомендацією продукту, а не пропозиція нової системи рекомендацій.

У (Musto et al., 2019) автори представляють підхід для створення обґрунтувань природної мови для підтримувати пропозиції, які повертає загальний алгоритм рекомендацій, заснований на обробці природної мови та методах аналізу настроїв для ідентифікації ввідних даних та ввідних аспектів, які характеризують елемент. Експериментальні результати, включені в документ, показують, що створення обґрунтування можуть бути багатими задоволенням, що користувачі надають перевагу на основі огляду обґрунтування на противагу іншим стратегіям пояснення (пояснення на основі ознак); зокрема, Пропозиції авторів порівнюються експериментально з ExplOD (Musto та ін., 2016). Як останній приклад, а Система рекомендацій, яка використовує аналіз тексту на думках, написаних китайською мовою, представлена в (Miao and Ланг, 2017). Матриця характеристик предмета будується для обчислення подібності характеристик предмета з метою щоб підвищити точність схожості предметів у тому, як рекомендації. Техніки під назвою ItemCF використовуються для створення рекомендацій (Shi et al., 2014), який підтримує попередній розрахунок подібних предметів.

Існують інші роботи, в яких використовуються числові оцінки, надані користувачами, текстові відгуки користувачів, на основі гіпотези, що це спільне використання може призвести до більш точних рекомендацій. Таким чином, у (Якоб та ін., 2009), який зосереджується на сфері фільмів,

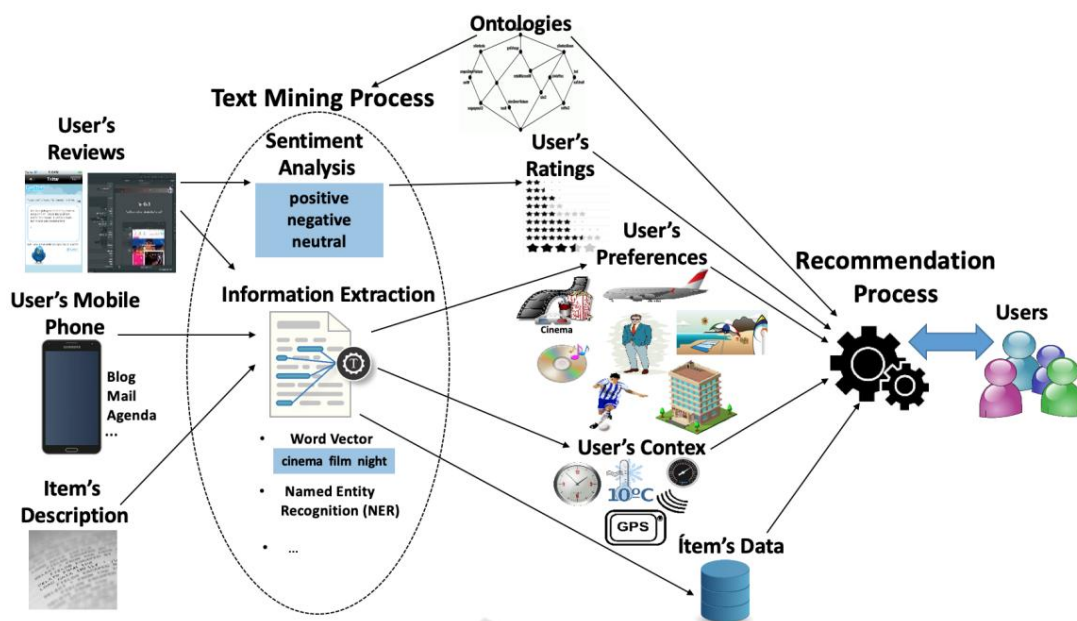


Рисунок 1. Використання аналізу тексту в системах рекомендацій: загальна структура.

вилучення думки на рівні фрази з фільму з довільним текстом перекладаються виконуються для виявлення тверджень з позитивними та негативними думками, ця інформація включається в систему рекомендацій під назвою HYRES (HYbrid REcommendation System). Крім того, фільми визначаються кластери аспектів в обчислюються глобальні числові значення для кожного кластера. Експериментальні результати показують, що прихований розподіл Др-хле (LDA) підходить до кластеризації даних найкращі результати.

Відгуки користувачів також можуть бути проаналізовані для вилучення контекстної інформації. Наприклад, (Li et al., 2010), яка фокусується на відгуках про ресторани, використовує текст методи інтелектуального аналізу для отримання інформації про 4 ризик-атрибути контексту: компаньйон (незалежно від того, чи є користувач один або з іншими людьми), привід для подорожі (наприклад, чи це свято, чи день народження), час дня та місця знаходження (у цьому випадку взято з профілю користувача, оскільки автори стверджують, що він зазвичай в дусі у текстах рецензій). Потім автори пропонують мовний латентну реляційну модель (PLRM) для інтеграції контекстних даних надання рекомендацій.

3 РЕКОМЕНДАЦІЙНІ СИСТЕМИ

ВИКОРИСТАННЯ ТЕКСТОВОГО МАЙНІНГУ НА ІНШІ ТЕКСТИ КОРИСТУВАЧА

У цьому розділі ми розглядаємо роботи, які вивчають текстові дані, написані користувачем (крім відгуків про товари, які вже були проаналізовані в розділі 2).

щоб отримати корисну інформацію для процесу рекомендацій (див. таблицю 2). Це може означати аналіз тексту для отримання інформації про контекст користувача, наприклад його/її місця знаходження (Др-монас Пфозер, 2010 рік; Feldman et al., 2015) та/або інформацію про його/її уподобання (наприклад, теми, якими він/вона цікавиться). Вхідними текстовими даними можуть бути, наприклад, твіти, написані користувачем. Крім того, багато останніх досліджень почали досліджувати інші джерела інформації, такі як соціальні мережі, де більшій текст повідомлення можна публікувати та ділитися. Щоденник користувача може бути іншим важливим джерелом тексту. Нарешті, кілька недавніх досліджень також продемонстрували ефективність використання колективних знань для збагачення профілю в інтересах користувача (Faralli та ін., 2015).

Twitter може бути цікавим джерелом даних, яке варто спробувати зробити висновки про інтереси користувачів (Piao and Breslin, 2018). Як приклад (Xu et al., 2011), починається з даних, що користувач часто публікує галасливих повідомлення про своє життя або створює розмови з друзями, які насправді не стосуються тем, що їх цікавлять. Тому вони запропонували структуру для вирішення цієї проблеми шляхом впровадження модифікованої моделі автор-суб'єкт під назвою Модель користувача Twitter, яка є генеративною моделлю розширює LDA, щоб включити інформацію про авторство. Прихована зміна використовується, щоб вказати, чи пов'язаний твіт з інтересом автора. На думку авторів, ця Модель можна розглядати як початкову роботу для багатьох завдань в Twitter, так як рекомендації друзів, оцінювання користувачів в аналізі соціальних мереж. Як інший наприклад, у пропозиції, представлений у (Zarrinkalam та ін., 2015) кожна цікава тема є поєднанням

Таблиця 1: Приклади систем рекомендацій, які використовують текстовий аналіз в дугу користувачів.

Предмети	Використані дані	Рекомендації Підхід	Підтримуючі інструменти	Методи аналізу тексту	Довідка
Продукти	Відгуки	Спільна фільтрація	Онтологія, так показники, як OFQ (Загальна якість функцій)	На основ правил методи класифікації, TMSK (Текстовий Майнер <small>комплекс програмного забезпечення</small>), програми забезпечення Riktext, онтологія (якість думки та якість продукції).	(Acliar та ін., 2007)
Предмети	Відгуки (написано в китайській)	Спільна фільтрація	Евклідова відстань, ItemCF (Штанн, 2014)	Сегментація слів (jieba), видалення стоп- слів, шаблони POS, HowNet (емоційний словник), кластеризація, статистика х-квадрат, логістична регресія	(Мяо та Ланг, 2017)
Туристична пам'ятки	Відгуки	На основ знань	Контрольовані алгоритми Проб-т-алгоритм, бномальна логістична регресія, дерева рішень (CHAID, C&RT Випадковий ліс)	IBM SPSS Modeler Аналітика тексту, видалення нелінійних об'єктів	(Геррейро та Ріта, 2019)
Покупки	Відгуки	TRS LoI (підхід на основі рівня довірливості)	LoT (рівень довіри)	Видалення стоп-слова, індексація слів, порівняння слів	(Абуейн та ін., 2017)
Фільми	Відгуки, огляди	Спільна фільтрація	Рекомендатор HYRES	Кластеризація, зупинка видалення слів, Явний семантичний аналіз (ESA), кластеризація на основі LDA, лексикон суб'єктивних підказок	(Якоб та ін., 2009)
Ресторани	Ресторан огляд	АВТОМОБІЛІ (рекомендація на основ контексту інформації)	Булева модель, Імовірнісний латентний Реляційна модель (PLRM)	Сумка слів, класифікатори на основі правил, Інструменти GATE (ANNIE, яка називається часом розпізнавання сутності)	(Лі та ін., 2010)

Кількість понять, які тимчасово спливають в Twitter. На основ цього виділяються активні теми протягом заданого інтервалу часу та схильність користувача до тих тем визначено. На основ цієї дії будується концептуальний граф застосовуються методи виявлення спільноти для виявлення активних тем, що цікавлять, у заданому інтервалі часу, враховуючи, що зв'язки між двома темами в соціальній мережі можуть змінюватися з часом. Крім того, запропоновано методику визначення позиції даного користувача щодо активних тем. Представлена пропозиція була застосована для розробки персоналізованої системи рекомендацій новин, яка була оцінена експериментально.

У зв'язку з наявним переважанням інформацією в Twitter, також були створені системи рекомендацій, щоб рекомендувати корисні теми, якими користувач могли б справді зацікавитися. Наприклад, у (Chen et al., 2012) для рекомендацій розглядаються наступні елементи: фактори, пов'язані з тематичним рівнем теми (використовується для виявлення спільних інтересів користувачів у певній темі), чинники соціального зв'язку

зв'язки користувача та явні характеристики, так як авторитет редактора та якість теми. Результати експерименту показують, що поєднання всіх цих елементів може допомогти покращити ефективність рекомендаційної теми.

Крім того, ще одним популярним типом даних, написаних користувачами, можуть бути текстові повідомлення, опубліковані в соціальних мережах. KBRS (Rosa et al., 2019) — це система рекомендацій на основі знань, яка на основі онтології та аналізу почуттів здатна аналізувати речення, опубліковані в соціальних мережах онлайн, з метою виявлення користувачів з потенційними психологічними розладами. розлади (депресія стрес). Потім, якщо необхідно, система рекомендацій використовується для надсилання повідомлень про щастя, спокій, розслаблення чи мотивацію. Текстові речення аналізуються за допомогою алгоритму машинного навчання, включаючи модель CNN, BLSTM-RNN, з урахуванням метрики відчуття (eSM2). eSM2 було змодельовано для покращення продуктивності KBRS, враховуючи параметри профілю користувача, його/її географічне розташування

Таблиця 2: Приклади систем рекомендацій, які використовують аналіз текстів в інших користувачів.

Предмети	Використані дані	Рекомендація	Підтримуючі інструменти	Видобуток тексту техніки	Довідка
Твітти	Твітти, користувачі соціальних даних, та інші особливості	Спільна фільмів	Модель факторизації, Подвійна Жаккарда	Приховані фактори	(Chen et al., 2012)
Повідомлення	повідомлення, профіль користувача, параметри, географічний розташування	На основі знань	глибоке навчання, рекурсивний нервовий мереж (РНН), Модель БЛСТМ-РНН, онтологія Нуаду	Алгоритми машинного навчання, метрика настрою (eSM2)	(Розата, 2019)
Туризм ROI	Текстовий повідомлення в веб-чаті	На основі вмісту фільмування	Імовірнісні методики (нечіткі мислення) онтології	Алгоритми Рокко і та Найвного Байєса, онтології туризму	(Loh et al., 2003)
Теги	Текст у тегах	Спільна фільмів	Асоціація правила, LDA	TF-IDF	(Krestel et al., 2009)

визначити інтенсивність почуття, визначеного в повідомлення. У роботі показано ефективність використання онтології та персоналізованого аналізу почуттів.

Техніки аналізу тексту також можна використовувати, щоб виявити інтереси користувачів у тексті, яким обмінюються в чаті. Наприклад, (Loh et al., 2003) експлуатували повідомлення в чатах між клієнтами турагентом, щоб з'ясувати та рекомендувати варіанти подорожей для клієнтів, особливо для тих, хто не знає куди йти, що там робити. Пропонована система запитує в онтології туризму для визначення ключових тем у тексті повідомлення, а потім запитує базу даних для отримання в подальших туристичних варіантах (такі як міста та пам'ятки). У цьому випадку система рекомендацій прихована від клієнта, оскільки турагент є тим, хто отримує рекомендації, які його/її підтримують, щоб краще направляти клієнта. Замість використання екстенсивної форми з параметрами, атрибутами та вимогами потреби та бажання клієнта виникають під час природної розмови та даних, отриманих системою, може доповнити потенційну дискусію досвід або знання турагента щодо спеціальних інтересів замовника.

Також була проведена певна робота з отримання інформації з текстів, написаних користувачами в неформальних комунікаціях (миттєві повідомлення, записи, зустрічі, протоколи, електронні листи) як потенційне джерело знань всередині організації. Це може бути дуже корисним, наприклад, для створення профілів користувачів для використання в системах рекомендацій. У (Gentile et al., 2011) автори стверджують, що можна змоделювати досвід людей автоматично шляхом відстеження неформальних комунікаційних обмінів (електронні листи) через семантичну анотацію їхнього вмісту для отримання динамічного користувача профілю. Потім профіль використовується для оцінки схожості між людьми (за допомогою індексу Жаккара). Інструмент під назвою SimNET (Similarity and Network Exploration Tool) інтерактивно відображає вміст користувача

мережі як частину можливостей пошуку та навігації, яка надає систему управління знаннями. Для створення профілів використовуються три техніки: Профіль на основі ключових слів, використовуючи Java Automatic Набір інструментів для розпізнавання термінів (JATR v1.02) для ключових слів видобуток; Профіль на основі Entity, використовуючи Open Веб-служба Calais для вилучення іменований об'єктів; профіль на основі концепції, використовуючи веб-сервіс Wikify для вилучення концепцій.

Інший актуальний тип письмового тексту, який може бути розглядається текст, присутній у тегах. Системи тегування стали важливими для Інтернету, оскільки вони дозволяють користувачам створювати теги, які анотують класифікують різні типи вмісту, ділитися ними з іншими. Однак важливою складністю є те, що теги є не обмеженим контрольованим словником анотацій, настанови, вони, як правило, шумні та рідкісні. The метапідходу, представленого в (Krestel et al., 2009) полягає в тому, щоб подолати проблему холодного запуску для тегування нових ресурсів. Зокрема, LDA використовується для отримання латентних тем з текстів, на основі цього, інші теги також можна рекомендувати належним чином до цих тем. В подальшому до представленої експериментальної оцінки, цей підхід забезпечує кращу продуктивність (з точки зору точності та запам'ятовування), ніж альтернативний підхід за допомогою правил асоціації.

4 РЕКОМЕНДАЦІЙНІ СИСТЕМИ ВИКОРИСТАННЯ ТЕКСТОВОГО МАЙНІНГУ НА ДАНІ ПРО ТОВАРИ

У цьому розділі ми розглядаємо деякі роботи, які застосовують текст аналіз даних, пов'язаних з елементами, які можна рекомендувати (див. таблицю 3).

З геопросторовими геотегами об'єктів, які мають як географічне розташування, так текстовий де-

Таблиця 3: Приклади систем рекомендацій, які використовують текстовий аналіз даних елементів.

Предмети	Використані дані	Рекомендаційна підсистема	Підтримуючі інструменти	Методи аналізу тексту	Довідка
Туризм POI	Ключові слова, GPS розташування, Описи POI	ReCoSKQ	CoSKQ	Будь-яка техніка видобутку тексту	(Hermoso et al., 2019)
Програми	Метадані версії, текстові описи	Гбрид	Напівконтрольована тематична модель, PMF, LDA	Розподіл топ-к, в, латентні топ-ки	(Лін та ін., 2014)
Програми	Опис контексту, назва та опис програми, метадані програми	ІЧ-підхід (контекстно-залежний).	пошукова система Lucene, Контекстно-залежний браузер (Копола та ін., 2010)	TF-IDF, кластеризація	(Міццаро та ін., 2014)
Події	GPS розташування, новини	Розташування	Усунення неоднозначності та вдповдність моделі навчання	СЛДА, СВМ (полном, гаусівський, лінній)	(Хотан, 2012)

scription, так звані запити з просторовими ключовими словами, які використовують як розташування, так і текстові описи, привертають все більше увагу (Cao et al., 2011). Більш конкретно, метою колективного запиту просторового ключового слова (CoSKQ) є отримання з просторової бази даних групи просторових елементів таким чином, щоб опис елементів, включених до цього набору, на основі ключових слів був повністю охоплений ключовими словами запиту, а також те, що отримані елементи знаходяться якомога ближче до місця розташування запиту та мають найменшу можливу відстань між самими елементами. У (Hermoso et al., 2019) автори стверджують, що використання цієї концепції може бути корисним для розробки систем рекомендацій для туризму, щоб надати користувачеві набір точок інтересу (POI), які задовольняють його/її запит (наприклад, запит, що виражає його/її поточний інтерес або потребу) як географічно, так і семантично; представлена для Re-CoSKQ, яка є адаптацією CoSKQ для сфери рекомендаційних систем. Хоча методи аналізу тексту явно не застосовуються в Re-CoSKQ, очікується, що вони знадобляться як етап попередньої обробки, щоб отримати ключові слова, які описують рідні елементи та/або профіль користувача.

Система рекомендацій мобільних додатків, яка використовує опис додатка та пов'язані з ним метадані, представлена в (Lin et al., 2014). Оскільки оновлена версія програми може внести значні зміни, вона спочатку генерує приховані теми з функцій версії, використовуючи напівконтрольовану модель тем для характеристики кожної версії. Темі розрзняються на основі метаданих пропонується техніка рекомендацій, яка називається VSR (рекомендація з урахуванням версії). Для експериментальної оцінки два альтернативних рекомендаційних підходи вважаються базовими: факторизація матриць ймовірностей (PMF), яка є технікою спільної фільтрації (CF), LDA, яка є фільтрацією на основі вмісту

(CBF) техніка. Автори показують точність рекомендацій, отриманих як при самостійному використанні цих методик, так і при їх комбінуванні (CF+CBF, CF+VSR, CBF+VSR та CF+CBF+VSR). У цьому роботі встановлюється, що гібридна система рекомендацій, яка включає запропоновану модель, чутливу до версії, досягає кращих результатів.

Інші системи рекомендацій мобільних програм також враховують контекст користувача. Таким чином, App-CAB (Mizzaro et al., 2014) є рекомендованою системою мобільних додатків, яка забезпечує проактивну та повністю автоматизовану процедуру запиту до ринку мобільних додатків, здатну отримувати набір додатків в класифікувати їх в відповідно до поточної ситуації користувача. Опис контексту користувача генерується контекстно-залежним браузером (Копола та ін., 2010) для пошуку в відповідних програмах для потреб користувача в цьому конкретному контексті. Щоб класифікувати контексти, усі слова в назві та описі додатків, які мають ту саму категорію, групуються, утворюючи кілька наборів слів, які представляють категорії ринку. Процес індексування слів виконується за допомогою Lucene для створення індексу категорій. Коли система аналізує поточний контекст, вона перевіряє кожне слово, яке описує контекст в індекс, щоб обчислити оцінки релевантності та отримати список потенційно релевантних категорій. Таким чином можна зважити доречність різних типів програм для даного контексту. Крім того, якщо слово, яке зустрічається в контексті, є частиною назви програми, то до початкового балу додається +10%, щоб підкреслити важливість збігу на рівні назви програми. Нарешті, застосовано крок фільтрації для усунення повторюваних результатів перед представленням їх користувачам.

Нарешті, в (Ho et al., 2012) представлений підхід до вилучення майбутніх просторово-часових подій з Інтернету для використання як кандидатів у системі рекомендацій щодо розташування подій. Ця робота пропонує

процедура виділення подій з новин статей, який складається з двох основних кроків: розпізнавання та зіставлення. У стадії розпізнавання місця мена та шаблони майбутнього часу ідентифікуються та екстрагуються. На етапі зіставлення виконуються операції просторово-часового усунення неоднозначності, дедуплікації та зіставлення. Змінна настрою (позитивна, негативний або нейтральний) додається до кожної події як аподомога для оформлення рекомендації. Отже, ідентифікована майбутня подія складається з її географічного розташування, часової моделі, змінної настрою, назви новини, ключової фрази та URL-адреси новинної статті.

5 ВИСНОВКИ

У цій статті ми надали огляд робіт з використання методів інтелектуального аналізу тексту в системі ges-omptender, характеризуючи їх в двох аспектах: тип аналізу зованих текстових даних (в дуги користувачів, інші тексти, написані користувачем, або асоційовані текстові дані до предметів) та з урахуванням їх призначення (витяг інформації про контекст користувача або виконання профілювання користувача).

Крім огляду в користувачів, використання різних типів текстів, написаних користувачем, здаються областю, яка залишається досить недослідженою, особливо щодо особистих текстових даних, так як отримані або надіслані електронні листи, особистий розпорядок дня користувача або щоденник користувача; це може бути частково через проблеми конфіденційності. Крім того, більшість робіт зосереджено лише на певному типі тексту, як у майбутніх робот ми маємо намір запропонувати інтегрований підхід, який поєднує та адаптує кілька технік як допоміжний інструмент для побудови контекстно-спеціалізованого рекомендаційного оцінювача та оцінювача його роботи.

ПОДЯКА

Ця робота виконана за підтримки проекту TIN2016-78011-C4-3-R (AEI/FEDER, UE) DGA-FSE (дослідницька група COSMOS).

СПИСОК ЛІТЕРАТУРИ

Abuein, Q., Shatnawi, A., and Al-Sheyab, H. (2017). Довірена система рекомендацій на основі рвня довір (TRS LoT). У 2017 році Міжнародна конференція з інженерії та технологій (ICET), сторінки 1–5.

Абуейн, К., Шатнав, М., Бан-Яссейн, М., Батха, Р. (2016). Структура для соціальних медіа та аналізу текстового контенту для цільового управління подіями. Міжнародний огляд комп'ютерів в програмного забезпечення, 11:388–394.

Aciar, S., Zhang, D., Simoff, S., and Debenham, J. (2007). Інформований рекомендацій: ґрунтуючись на рекомендаціях огляди споживчих товарів в Інтелектуальних системах IEEE, 22(3):39–47.

Адомавчус, Г., Тужилін, А. (2011). Рекомендатор Довідник з систем, розділ Контекстно-залежних систем рекомендацій, сторінки 217–253. Springer США, Бостон, Массачусетс.

Беррі, МВ, редактор (2004). Огляд інтелектуального аналізу тексту. Спрингер Нью-Йорк.

Сао, Х., Конг, Г., Jensen, СS, та Оої, BC (2011). Колективний пошук просторових ключових слів. У ACM SIGMOD Міжнародна конференція з менеджменту Дані (SIGMOD 2011), сторінки 373–384, Нью-Йорк, Нью-Йорк, США. ACM.

Чень, К., Чень, Т., Чжен, Г., Цзень, О., Яо, Е. та Юй, Ю. (2012). Спільний персонал зованих рекомендацій в Інтернеті. На 35-й Міжнародній конференції ACM SIGIR з досліджень розробок в галузі інформації Пошук (SIGIR 2012), сторінки 661–670. ACM.

Коппола П., Делла Меа В., Ді Гасперо Л., Менегон Д. М'єскіс, Д., М'єццаро, С., Ск'єньєтто, І., Вассена, Л. (2010). Контекстно-залежний браузер. IEEE Intelli-gent Systems, 25(1):38–47.

Др'ємонас, Е. та П'єозер, Д. (2010). Витяг геопросторового маршруту з тексту в У Першому міжнародному семінарі ACM SIGSPATIAL з аналізу даних для геоінформатики (DMG 2010), сторінки 29–37. ACM.

Фаралл С., Ст'єло Г. та Велард П. (2015). Рекомендації користувачів в м'єкроблогів на основі єрархічного інтересу профілю. Аналіз соціальних мереж майбутнього, 5(1).

Фельдман Д., Сунг К., Сугая А. та Рус Д. (2015). Щоденник: від сигналу в GPS до щоденника з текстовим пошуком. ACM Transactions on Sensor Networks, 11(4):60:1–60:41.

Джентиле, А.Л., Ланфранчі, В., Мазумдар, С., С'єравенья, Ф. (2011). Вилучення семантичних мереж користувачів з неформальних комунікаційних об'єктів в Аройо, Л., Велт, К., Алан, Х., Тейлор, Дж., Бернштейн, А., Кагал, Л., Ной, Н., Блумквіст, Е., редактори, The Semantic Web – ISWC 2011, сторінки 209–224. Спрингер.

Gruber, TR та Н. (1993). Підхід до перекладу специфікацій портативної онтології. Придбання знань, 5(2):199–221.

Guerreiro, J. та Rita, P. (2019). Як передбачити явні рекомендації в онлайн-оглядах за допомогою аналізу тексту та настроїв. журнал о Менеджмент гостинності та туризму. DOI: 10.1016/j.jhtm.2019.07.001. У пресі, виправлено доказ.

Гупта В., Лехал Г. С. та Н. (2009). Огляд методів застосування копіювання тексту. Journal of Emerging Technologies in Web Intelligence, 1(1):60–76.

Хегде, Ю. та Падма, Словаччина (2017). Аналіз настроїв використання випадкового л'єсового ансамблю для повторного перегляду мобільних продуктів у каналі. У 2017 році IEEE 7th International Advance Computing Conference (IACC), сторінки 777–782.

- Ермосо, Р., Ларр, С., Трелло-Ладос, Р. (2019). Re-CoSKQ: на шляху до рекомендаційної POI за допомогою запитів в колективних просторах ключових слів. На робочому семінарі ACM RecSys щодо рекомендацій у туризмі (RecTour 2019), у поєднанні з 13-ю конференцією ACM щодо систем рекомендацій (RecSys 2019), Копенгаген (Данія), том 2435, сторінки 42–45. Семінар CEUR праць, ISSN 1613-0073.
- Хо, С.-С., Лібман, М., Ван, П., Самет, Х. (2012). Дослідження майбутніх просторово-часових подій та їхніх настрів з статей онлайн-новин для системи рекомендацій з урахуванням розташування. На першому міжнародному семінарі ACM SIGSPATIAL з мобільних географічних інформаційних систем (MobiGIS 2012), сторінки 25–32, Нью-Йорк, Нью-Йорк, США. ACM.
- Якоб, Н., Вебер, С.Х., Мюллер, М.Ц., Гуревич, І. (2009). Поза межами зрок експлуатації користувача в логічному тексті огляди для підвищення точності рекомендацій щодо фільмів. Перший міжнародний семінар CIKM з тематичного аналізу настроїв для масової думки (TSA 2009), сторінки 57–64, Нью-Йорк, Нью-Йорк, США. ACM.
- Кім, Е.-Г., Чун, С.-Х. (2019). Аналіз автомобільного онлайн-огляди за допомогою аналізу тексту. Стаття, 11(6):1611.
- Крестел, Р., Фанкхаузер, П., Нейдл, В. (2009). Латентний Розподіл Діагностичних рекомендацій і тем. У Третій Конференції ACM щодо систем рекомендацій (RecSys 2009), сторінки 61–68. ACM.
- Ліу, Н. Дж., Чжан Ю., Ван Б., Янь Б. та Вен Ф. (2010). Контекстна рекомендація на основі тексту видобутку корисних копалин. На 23-й Міжнародній конференції з комп'ютерної лінгвістики (COLING 2010): постери, сторінки 692–700. Асоціація комп'ютерної лінгвістики.
- Ліу, Дж., Сугіяма, К., Кан, М.-Й., Чуа, Т.-С. (2014). Нове та покращене: версія моделювання для покращення рекомендаційної програми. На 37-й Міжнародній конференції ACM SIGIR з досліджень розробок у пошуку інформації (SIGIR 2014), сторінки 647–656, Нью-Йорк, Нью-Йорк, США. ACM.
- Лю Б. та Чжан Л. (2012). Опитування опитування Sentiment Analysis, сторінки 415–463. Springer США, Бостон, Массачусетс.
- Ло, С., Лоренц, Ф., Салдана, Р., Літню, Д. (2003). Система туристичних рекомендацій, заснована на співпраці та аналізі тексту. Інформаційні технології та Туризм, 6(3):157–165.
- Мяо, Д. та Ланг, Ф. (2017). Система рекомендацій на основі видобутку тексту. У 2017 році Міжнародна конференція з кберрозподілення обчислень Knowledge Discovery (CyberC), сторінки 318–321.
- Міццаро, С., Паван, М., Скандетто, І., Занелло, І. (2014). Контекстно-залежна система пошуку для мобільних додатків. На 4-му семінарі з усвідомлення контексту в пошуку та рекомендаціях (CARR 2014), сторінки 18–25. ACM.
- Мусто К., Лопс П., де Геммс М. та Семераро Г. (2019). Обґрунтування рекомендацій за допомогою аспектного аналізу настроїв в дугу користувачів. У 27-му Конференції ACM з моделювання користувачів, адаптації та Персоналізації, UMAP '19, сторінки 4–12, Нью-Йорк, Нью-Йорк, США. ACM.
- Musto, C., Narducci, F., Lops, P., De Gemmis, M., and Semeraro, G. (2016). ExpLOD: структура для пояснення рекомендацій на основі Linked Open Хмара даних. На 10-й конференції ACM щодо систем рекомендацій (RecSys 2016), сторінки 151–154. ACM.
- Piao, G. та Breslin, JG (2018). Визначення інтересів користувачів у микроблогах соціальних мереж: опитування. Моделювання користувача та адаптована до користувача взаємодія, 28(3):277–329.
- Річчі, Ф., Рокач, Л., Шапра, Б. (2011). Рекомендатор Systems Handbook, розділ «Вступ до Recommender Systems Handbook», сторінки 1–35. Springer США, Бостон, Массачусетс.
- Річчі, Ф., Рокач, Л., Шапра, Б. (2015). Рекомендатор Довідник з систем, розділ Системи рекомендацій: Вступ викиди, сторінки 1–34. Спрингер США, Бостон, Массачусетс.
- Роза Р. Л., Шварц Г. М., Руджеро В. В. та Родригес, DZ (2019). Система рекомендацій на основі знань, яка включає аналіз настроїв глибоке навчання. IEEE Transactions on Industrial Informatics, 15(4):2124–2135.
- Роул, Р. К. та Ароа, К. (2019). Чудовий огляд до системи рекомендацій на основі резюме тексту для електронних виробів. Soft Computing, 23(24):13183–13204.
- Shi, Y., Larson, M., and Hanjalic, A. (2014). Спільний фільм за межами матриці елементів користувача: огляд сучасного стану майбутніх викидів. Обчислення ACM Опитування, 47(1).
- Субраманьясам В., Маногаран Г., Логеш Р., Ваджайкумар В., Чламкурт Н., Малат Д. та Senthilselvan, N. (2019). Онтологічно орієнтований подвійний рекомендаційний щодо іж в охороні здоров'я на основі Інтернету речей система. Журнал суперкомп'ютерів, 75(6):3184–3216.
- Тарус, Дж.К., Нгу, З., Мустафа, Г. (2018). Рекомендація на основі знань: огляд на основі онтології рекомендаційних системи для електронного навчання. Огляд штучного інтелекту, 50(1):21–48.
- Xu, Z., Ru, L., Xiang, L., and Yang, Q. (2011). Виявлення інтересу користувачів у Twitter за допомогою модифікованої моделі автор-тема. У 2011 році IEEE/WIC/ACM International Conference з Web Intelligence та Intelligent Agent Технологія, том 1, сторінки 422–429.
- Заррнкалам, Ф., Фан, Х., Багер, Е., Кахан, М., Ду, В. (2015). Виявлення інтересів користувачів підтримкою семантики з твітера. У 2015 році IEEE/WIC/ACM International Conference з веб-розвідки та інтелектуального агента Технологія (WI-IAT), том 1, сторінки 469–476.