

ОСОБЛИВОСТІ РЕКОМЕНДАЦІЙНОГО АЛГОРИТМУ НА ОСНОВІ АНАЛІЗУ МЕТОДІВ ДОБУВАННЯ ДАНИХ З СОЦІАЛЬНИХ МЕРЕЖ

Петро Смачило¹, Любов Журавчак²

Національний університет “Львівська політехніка”,
кафедра програмного забезпечення, вул. С. Бандери, 12, Львів, Україна

¹ petro.i.smachylo@lpnu.ua, ORCID 0009-0008-1285-4181,

² liubov.m.zhuravchak@lpnu.ua, ORCID 0000-0002-1444-5882

© Смачило П., Журавчак Л., 2023

В останні роки платформи соціальних мереж стали потужними інструментами зі збирання даних для покращення досвіду користувачів. Величезна кількість даних, отриманих через соціальні медіа, надає унікальну можливість для розроблення інноваційних систем рекомендацій. У статті проаналізовано застосування методів інтелектуального аналізу даних щодо соціальних мереж у контексті ефективних систем рекомендацій, зосереджено увагу на трьох ключових методологіях: аналізі тональності тексту (АТТ), тематичному моделюванні (ТМ) й аналізі соціальних мереж (АСМ), виокремлено їхні позитивні риси.

АТТ дає змогу системі адаптувати рекомендації на основі аналізу настроїв, пропонуючи користувачам предмети, які відповідають їхнім вираженим емоціям. Експерименти показують істотне підвищення точності рекомендацій, коли дані настрою інтегровані. ТМ дає змогу системі зрозуміти основні проблеми користувачів, визначаючи провідні теми, надаючи індивідуальні рекомендації та залишаючись у курсі тенденцій, що розвиваються. Водночас АСМ визначає впливових користувачів і спільноти, підвищуючи релевантність і обізнаність про елементи системи.

У статті підкреслено величезний потенціал соціальних мереж для розроблення ефективних, персоналізованих систем рекомендацій. Використовуючи аналіз настроїв та тематичне моделювання, ці системи можуть надавати персоналізовані та релевантні рекомендації на основі суспільних настроїв, популярних тем і динаміки соціальних мереж.

Ключові слова: аналіз даних; соціальні мережі; система рекомендацій; аналіз тональності мови; тематичне моделювання.

1. Постановка задачі та формулювання мети статті

Зростання популярності платформ соціальних медіа революціонізувало спосіб спілкування людей та їхньої взаємодії [1]. Окрім особистого спілкування, соціальні мережі стали центрами поширення інформації. У цій статті дослідимо потенціал використання методів інтелектуального аналізу даних для розроблення надійної системи рекомендацій внаслідок опрацювання великої кількості даних, отриманих через соціальні мережі.

Дедалі більше використання соціальних медіа з метою створення рекомендаційних систем поряд з відкриттям нових можливостей змушує дослідників також шукати відповіді на виклики, що

постають перед ними. Незважаючи на велику кількість наявних даних, ефективне їх опрацювання для надання рекомендацій потребує вирішення таких завдань:

1. Зашумленість і неформальність даних. Дані соціальних медіа часто зашумлені, неофіційні та містять широкий спектр мовних варіацій, що ускладнює отримання значущої інформації.

2. Занепокоєння щодо збереження конфіденційності. Забезпечення анонімності та належного захисту особистих даних користувачів соціальних мереж після їх збирання та аналізу має вирішальне значення.

3. Динамічність соціальних медіа. Платформи соціальних медіа є динамічними, тенденції та дискусії постійно змінюються, тому система повинна адаптуватися до цих змін і надавати рекомендації в реальному часі.

4. Алгоритмічні упередження. Алгоритми інтелектуального аналізу даних можуть ненавмисно вводити упередження на основі даних, на яких вони навчаються, що потенційно може призвести до неправильних рекомендацій.

Вирішення цих завдань і досягнення окреслених цілей сприятиме розробленню ефективної та етично обгрунтованої системи рекомендацій на основі аналізу даних із соціальних мереж.

Метою дослідження є аналіз наявних методів видобування даних задля подальшого їх упровадження у рекомендаційну систему та побудова алгоритму, який допоможе підвищити точність рекомендацій у системі.

2. Аналіз останніх досліджень та публікацій

Аналіз тональності тексту (АТТ, також відомий як аналіз думок) зосереджується на виявленні та аналізі настроїв, виражених у публікаціях у соціальних мережах. Це процес виявлення контекстуальної полярності тексту, визначення його позитивності, негативності чи нейтральності. Друга назва підходу відображає думку чи ставлення мовця [2]. Вивчення емоційного тону, ставлень та думок, висловлених у публікаціях, аналіз настроїв може надати цінну інформацію про сприйняття та настрої громадськості щодо конкретних благодійних організацій та філантропічних ініціатив.

В АТТ використовують різні методи для класифікації настроїв публікацій у соціальних мережах, зокрема алгоритми машинного навчання, методи опрацювання природної мови (NLP), підходи на основі лексиконів (словників настроїв) та графів.

Алгоритми машинного навчання можна навчити на позначених даних, щоб автоматично класифікувати публікації як позитивні, негативні чи нейтральні на основі мови та контекстуальних ознак, наявних у тексті. Машинне навчання – найперспективніший підхід для АТТ. Зазвичай АТТ на основі машинного навчання забезпечує вищу оцінку точності, ніж підхід на основі лексиконів. Він пропонує різні методи розроблення функцій, які виділяють критичні характеристики з набору даних і підвищують ефективність АТТ. Для класифікації настроїв доступні різні керовані та некеровані алгоритми. Керований підхід працює із позначеними наборами даних і використовує функцію зіставлення вхідних міток із вихідними мітками [3].

Методи NLP дають змогу аналізувати лінгвістичні особливості та шаблони в тексті, такі як слововживання, структура речення та синтаксичні шаблони, щоб визначити основне почуття.

Лексикони, які пов'язують слова з балами настроїв або полярністю, дають можливість класифікувати настрої на підставі наявності та інтенсивності конкретних слів. Такий підхід починається з попереднього опрацювання аналізованого тексту. Спершу загальний бал прирівнюють до нуля, потім перевіряють, чи є слово у словнику, і якщо так, то визначають його позитивність чи негативність, і оцінку, відповідно, оновлюють. Тому підсумкова оцінка класифікує текст як позитивний або негативний [4].

Підхід на основі графів [5] з'єднує взаємопов'язані слова в текстових оглядах, щоб визначити настрої та думки людей, тут вершини (вузли) відповідають характеристикам, доступним у оглядах.

Однак АТТ стикається з певними труднощами в контексті даних соціальних медіа. Пости в соціальних мережах часто короткі, неформальні та містять розмовні вирази, аббревіатури та смайли, що може ускладнити аналіз тональності тексту. Крім того, сарказм, іронія та культурні нюанси, наявні в тексті, можуть вплинути на точність класифікації настроїв [6]. Для вирішення цих проблем потрібно розробити надійні моделі, які зможуть працювати з неформальною мовою та враховувати контекстуальні підказки й тонкощі, притаманні комунікації в соціальних мережах.

Отже, АТТ відіграє вирішальну роль у видобуванні даних із соціальних мереж для рекомендацій. Завдяки точному визначенню та класифікації настроїв, виражених у публікаціях у соціальних мережах, він дає змогу надавати персоналізовані та цілеспрямовані рекомендації, які відповідають настроям і вподобанням користувачів. Він також надає цінну інформацію для організацій, допомагаючи їм зрозуміти суспільні настрої та вдосконалити свої стратегії.

Тематичне моделювання (ТМ) – це потужна техніка аналізу даних, спрямована на розкриття основних тематичних шаблонів і тем у колекції публікацій у соціальних мережах. Воно надає засоби для отримання значущої інформації з неструктурованих текстових даних і для класифікації дописів у послідовні тематичні кластери. Розроблено кілька підходів для вирішення проблеми тематичного моделювання, кожен із яких пропонує унікальне розуміння виявлення значущих тем у неструктурованих текстових даних.

Використовуючи алгоритми ТМ, такі як прихований розподіл Діріхле (Latent Dirichlet Allocation, LDA), система рекомендацій може визначати найрелевантніші та найпопулярніші благодійні кампанії, які обговорюють на платформах соціальних медіа, забезпечуючи персоналізовані та цільові рекомендації для потенційних донорів. ТМ ґрунтується на статистичному моделюванні спільного використання слів у тексті для висновку про приховані теми. Алгоритм LDA, зокрема, передбачає, що кожен документ у наборі даних є сумішшю тем і кожна тема характеризується розподілом ймовірностей за словами [7]. За допомогою ітераційного аналізу слів у документах і оцінювання їх розподілу за темами LDA виявляє основні теми та відповідний розподіл слів. Це може дати змогу ідентифікувати та класифікувати ключові теми, які обговорюють у публікаціях у соціальних мережах, пов'язаних із благодійністю.

Латентний семантичний аналіз (LSA) – метод тематичного моделювання в галузі опрацювання природної мови та аналізу тексту, привернув значну увагу та отримав схвальні відгуки за унікальну методологію та можливість застосування у різних сферах. LSA, оснований на декомпозиції сингулярного значення (SVD), призначений для виявлення прихованих семантичних структур, на яких основані часто складні та багатогранні зв'язки між термінами та документами в межах певного корпусу. Латентний семантичний аналіз є чудовим підходом у сфері тематичного моделювання. Базування на SVD, зменшення розмірності та відкриття прихованих семантичних структур робить його цінним інструментом для вилучення значення та розкриття основних тем у текстових даних. Починаючи з пошуку інформації та завершуючи кластеризацією документів і резюмуванням тексту, LSA довів свою ефективність у різноманітних завданнях аналізу тексту, пропонуючи цінну інформацію про приховані семантичні зв'язки між термінами та документами, зрештою, покращуючи розуміння та використання великих текстових корпусів [8].

Факторизація невід'ємної матриці (NMF) є переконливою альтернативою традиційним підходам тематичного моделювання, таким як LSA. NMF пропонує унікальні переваги, особливо в сценаріях, де інтерпретація є критичною вимогою, наприклад, аналіз тексту на основі вмісту та додатки, в яких перевагу віддають суто позитивним компонентам. За суттю NMF працює на передумові невід'ємності, важливій характеристиці, яка відрізняє його від LSA. NMF розділяє терм-документну матрицю на дві невід'ємні матриці: одна представляє зв'язки термін – тема, інша – фіксує

зв'язки тема – документ. Обмеження невід'ємності гарантує, що всі елементи в цих матрицях залишаються строго позитивними, що узгоджується з уявленням про те, що теми в тексті є адитивними, складаються зі слів, які позитивно впливають на їх представлення [9].

Однак ТМ стикається з певними труднощами у разі застосування до даних соціальних мереж. Пости в соціальних мережах часто характеризуються стислістю, неформальністю та використанням хештегів, емодзі та інших нетекстових елементів [10]. Ці характеристики можуть ускладнювати точне охоплення та моделювання основних тем. Крім того, платформи соціальних медіа є багатими джерелами мультимедійного контенту, урахуовуючи зображення та відео, які потребують додаткових методів ТМ, крім текстового аналізу.

Вирішення цих проблем потребує інтеграції передових методів, які можуть працювати з унікальними характеристиками даних із соціальних мереж. Методи, які передбачають мультимодальний аналіз, такі як поєднання розпізнавання зображень і аналізу тексту, можуть підвищити точність і повноту моделювання теми в контексті даних із соціальних мереж.

Отже, ТМ є важливим підходом до аналізу публікацій у соціальних мережах. Визначення та класифікація тем дають змогу системі рекомендацій пропонувати персоналізовані та відповідні пропозиції користувачам. Крім того, ТМ дає змогу виявляти нові тенденції та дискусії, підтримуючи систему в актуальному стані з урахуванням змін інтересів [11]. Оскільки дослідники продовжують удосконалювати методи ТМ, зокрема, мультимодальний аналіз, ефективність і застосовність ТМ щодо рекомендацій продовжуватимуть зростати.

Аналіз соціальних мереж (АСМ) – це потужний підхід до інтелектуального аналізу даних, який зосереджується на розумінні стосунків і взаємодії між користувачами в соціальних мережах. Відображаючи соціальні зв'язки та вплив окремих осіб, АСМ може дати цінну інформацію про динаміку та структуру цих мереж.

Одним із ключових аспектів АСМ є концепція показників центральності, які кількісно визначають важливість або вплив вузлів (окремих користувачів або об'єктів) у мережі на основі їхніх зв'язків та становища. Наприклад, ступінь центральності вузла вимірює кількість зв'язків, які він містить, тоді як центральність між вузлами означає доступність до інших вузлів у мережі [12]. Аналізуючи показники центральності, АСМ може ідентифікувати користувачів, які займають центральні позиції в соціальній мережі, вказуючи на їхній потенційний вплив і охоплення.

Іншим аспектом АСМ є виявлення спільноти. Спільнота в соціальних мережах означає групу вузлів, тісно взаємопов'язаних, які мають менше зв'язків із вузлами за її межами [13]. Алгоритми виявлення спільнот можуть ідентифікувати ці окремі групи в соціальних мережах.

Окрім виявлення впливових користувачів, АСМ також може надати уявлення про поширення інформації в соціальних мережах [14]. Аналізуючи моделі потоку інформації, він може виявити, як інформація про благодійні справи поширюється мережею та досягає потенційних жертводавців. Це розуміння може допомогти організаціям розробити ефективні стратегії поширення інформації та залучення соціальних зв'язків для максимізації впливу кампаній зі збирання коштів.

Однак важливо визнати, що АСМ має й обмеження. Аналіз залежить від наявності та якості даних, а на точність результатів впливають такі фактори, як повнота та надійність даних. Крім того, під час АСМ потрібно враховувати етичні принципи, зокрема, щодо конфіденційності та захисту даних.

Отже, АСМ – важливий підхід до аналізу даних для розуміння взаємозв'язків і динаміки впливу в соціальних мережах. Виявляючи впливових користувачів і спільноти, він може сприяти розробленню ефективних рекомендацій у системах громадського фінансування. Інтеграція АСМ з іншими методами інтелектуального аналізу даних, такими як АТТ і ТМ, може підвищити точність і релевантність рекомендацій (табл. 1). Оскільки АСМ продовжує розвиватися, він має великий потенціал для розширення засобів взаємодії благодійних організацій із потенційними донорами та збирання коштів для своїх проєктів.

Порівняння підходів до аналізу даних

	АТТ	ТМ	АСМ
Мета	Аналіз та класифікація почуттів, виражених у текстових даних	Розкриття основних тематичних шаблонів і тем у колекції публікацій у соціальних мережах	Вивчення відносин та взаємодії між користувачами в соціальній мережі
Підходи	Алгоритми машинного навчання, методи NLP, підходи на основі лексиконів	LDA, NMF, імовірнісний латентно-семантичний аналіз (PLSA)	Показники центральності, алгоритми виявлення спільнот, візуалізація мережі
Призначення	Розуміння та аналіз настрою щодо конкретних тем, продуктів або сутностей	Виявлення прихованих тематик, тем та зв'язків у колекції документів	Виявлення впливових користувачів, спільнот та потоків інформації у соціальній мережі
Застосування	Моніторинг репутації бренду, аналіз відгуків клієнтів, видобування думок	Організація контенту, кластеризація документів, пошук інформації, системи рекомендацій	Маркетинг впливу, вірусний маркетинг, аналіз соціального впливу, виявлення чуток
Виклики	Робота з неформальною мовою, сарказмом, іронією та культурними нюансами	Визначення оптимальної кількості тем, статистика тем, неоднозначність тем	Приватність даних, масштабованість, виявлення упередженості, зміни динаміки мережі
Переваги	Надає уявлення про громадську думку, настрої та ставлення до конкретних тем	Розкриває приховані тематичні структури та семантичні зв'язки в текстових даних	Виявляє впливових користувачів, спільноти та потоки інформації в соціальній мережі
Потенціал інтеграції	Може бути інтегрований з іншими методами для покращення рекомендацій	Може бути поєднаний з АТТ для покращення контекстно-орієнтованих рекомендацій	Може бути використаний разом з АТТ та ТМ для покращення рекомендацій

3. Особливості, які необхідно врахувати під час розроблення етичної та ефективної системи рекомендацій

Конфіденційність даних. Збирання та аналіз особистої інформації користувачів, видобування даних із соціальних медіа для отримання рекомендацій спричиняє серйозні занепокоєння щодо їхньої конфіденційності. Завдяки пріоритетності анонімізації користувачів та захисту даних можна забезпечити їхню конфіденційність, вселяючи довіру та впевненість у системі рекомендацій.

Якість і надійність даних. Дані соціальних медіа величезні, неструктуровані та часто зашумлені, що погіршує їхню якість та знижує надійність. Тому необхідно застосувати методи попереднього опрацювання та очищення даних для поліпшення їхньої якості. Треба виявити та видалити повторювані записи, відфільтрувати нерелевантні публікації, а відсутні або неповні дані належно опрацювати. Застосовуючи методології очищення даних, такі як виявлення та видалення дублікатів, нормалізацію тексту, можна підвищити точність і надійність рекомендацій, гарантуючи, що вони ґрунтуються на високоякісних даних.

Методи нормалізації тексту стандартизують текстові дані, роблячи їх узгодженими та придатними для аналізу. Нормалізація тексту відіграє ключову роль у поліпшенні якості та підвищенні ефективності аналізу даних соціальних мереж, стандартизуючи та очищуючи текстові дані [15]. Вона забезпечує узгодженість, перетворюючи весь текст на малі літери й запобігаючи впливу варіацій

великих літер на аналіз. Крім того, нормалізація тексту зменшує розмірність за допомогою таких методів, як коріння й лематизація, роблячи набір даних керованішим. Цей процес особливо цінний для завдань опрацювання природної мови, підвищуючи продуктивність аналізу настроїв, тематичного моделювання.

Динамічність вмісту соціальних мереж. Платформи соціальних медіа характеризуються динамічною природою, постійними оновленнями та зміною тенденцій. Щоб надавати релевантні рекомендації в режимі реального часу, система повинна ефективно адаптуватися до цих змін. Постійне збирання і моніторинг даних життєво важливі для фіксування останніх тенденцій, настроїв і дискусій, пов'язаних із благодійністю. Стежачи за динамічним вмістом соціальних мереж, система рекомендацій може надавати користувачам своєчасні та відповідні контексту пропозиції, максимізуючи їхнє залучення та потенційний внесок.

Алгоритмічні упередження. Алгоритми інтелектуального аналізу даних, які використовують в системах рекомендацій, можуть ненавмисно вводити упередження на підставі даних, на яких вони навчаються. Ці упередження можуть призвести до неправильних і спотворених рекомендацій [16]. Щоб забезпечити правильність і неупередженість, вкрай важливо вилучити алгоритмічні упередження, ретельно оцінюючи та перевіряючи дані навчання. Регулярні аудити та оцінки системи рекомендацій допомагають виявити та виправити будь-які упередження, які можуть виникнути. Пом'якшуючи алгоритмічні упередження, система рекомендацій може надавати рекомендації, які відповідають цінностям інклюзивності та правильності.

Урахування під час розроблення рекомендаційного алгоритму вказаних особливостей дасть змогу повністю розкрити потенціал розглянутих методів, забезпечити конфіденційність даних, покращити їхню якість і підвищити надійність, адаптуватися до динамічного вмісту соціальних мереж та пом'якшити алгоритмічні упередження. Система зможе надавати персоналізовані та етично обґрунтовані рекомендації користувачам, сприяючи більшій їхній залученості.

4. Особливості застосування різних методів добування даних із соціальних мереж у рекомендаційній системі

АТТ у контексті систем рекомендацій може бути надзвичайно цінним. Аналізуючи настрої, пов'язані з різними елементами, система рекомендацій може визначити причини, які позитивно резонують у користувачів. Наприклад, якщо користувач висловлює позитивні настрої щодо якоїсь теми у своїх публікаціях у соціальних мережах, система може рекомендувати елементи, пов'язані з цією темою. З іншого боку, якщо користувач демонструє негативне ставлення до певного виду діяльності, система може уникати рекомендацій елементів, пов'язаних із такою діяльністю, забезпечуючи персоналізованіший і відповідніший досвід для користувача.

Експерименти у дослідженні, де застосовували АТТ у рекомендаційній системі [17], показали, що оцінки користувачів щодо товарів не обов'язково співвідносяться з настроями відгуків, написаних для тих самих товарів. І навіть більше, системи рекомендацій можуть генерувати різні рекомендації щодо оцінки настроїв і тому по-різному працювати, коли йдеться про настрої в оглядах. Крім того, введення даних про настрої у процес гібридної рекомендації може істотно поліпшити її якість і, отже, запропонувати переваги користувачам. З набором даних цифрової музики, яку використовували автори, значення точності рекомендацій моделі без і з настроєм становили 2,2 % і 4,8 % відповідно (118 % покращення).

Крім того, АТТ може надати цінні відгуки та інформацію для громадських організацій. Відстежуючи та аналізуючи настрої, висловлені в соціальних мережах, ці організації можуть глибше зрозуміти громадське сприйняття та настрої щодо їхніх ініціатив. Ця інформація може допомогти їм удосконалити повідомлення, скоригувати стратегії та ефективніше взаємодіяти зі своєю аудиторією.

ТМ може відігравати ключову роль у розумінні основних проблем та інтересів користувачів. Визначаючи провідні теми та кластери в даних соціальних мереж, система рекомендацій може надавати індивідуальні пропозиції на основі конкретних інтересів користувачів. Наприклад, якщо користувач часто публікує дописи, пов'язані із збереженням навколишнього середовища, система

може рекомендувати благодійні організації, які займаються екологічною стійкістю або охороною дикої природи. Узгоджуючи рекомендації з тематичними вподобаннями користувачів, система покращує їхню взаємодію та збільшує ймовірність вагомих внесків.

Експерименти у дослідженні, де застосовували ТМ у рекомендаційній системі [18], показують, що поєднання колаборативного фільтрування – техніки, яка може відфільтрувати елементи, що можуть сподобатися користувачеві, на основі реакції схожих користувачів – з ТМ працює краще з погляду точності та запам'ятовування, зокрема, у сфері кіно, а його продуктивність є найвищою для п'яти тем у сфері ресторанів.

ТМ також може сприяти виявленню нових тенденцій і дискусій у будь-якій сфері. Відстежуючи еволюцію тем із часом, система рекомендацій може залишатися в курсі останніх проблем і причин, які набувають популярності на платформах соціальних мереж. Така обізнаність у режимі реального часу дає змогу системі надавати своєчасні та відповідні рекомендації, які відображають динамічність суспільних інтересів і проблем.

АСМ відіграє вирішальну роль у виявленні впливових користувачів, які можуть сприяти поширенню обізнаності про елементи системи та рекомендувати відповідні можливості своїм підписникам. У статті [19] запропоновано систему рекомендацій, яка застосовує аналіз соціальних мереж, із вищими точністю (98 %) і влучністю (91 %) порівняно з лише колаборативним фільтруванням. У запропонованій системі більше значення запам'ятовування (89 %) та оцінка F1 (комбінація точності та запам'ятовування із використанням середнього гармонічного, 90 %). Автори дійшли висновку, що запропонована система рекомендацій корисніша для користувачів щодо вибору кращих товарів.

Аналізуючи спільноти, АСМ може виявити кластери користувачів зі спільними інтересами та приналежністю. Цю інформацію можна використовувати, щоб рекомендувати елементи системи, які відповідають інтересам і вподобанням конкретних спільнот.

5. Архітектура системи

На рис. 1 подано архітектуру рекомендаційної системи щодо збирання коштів громадською благодійною організацією.

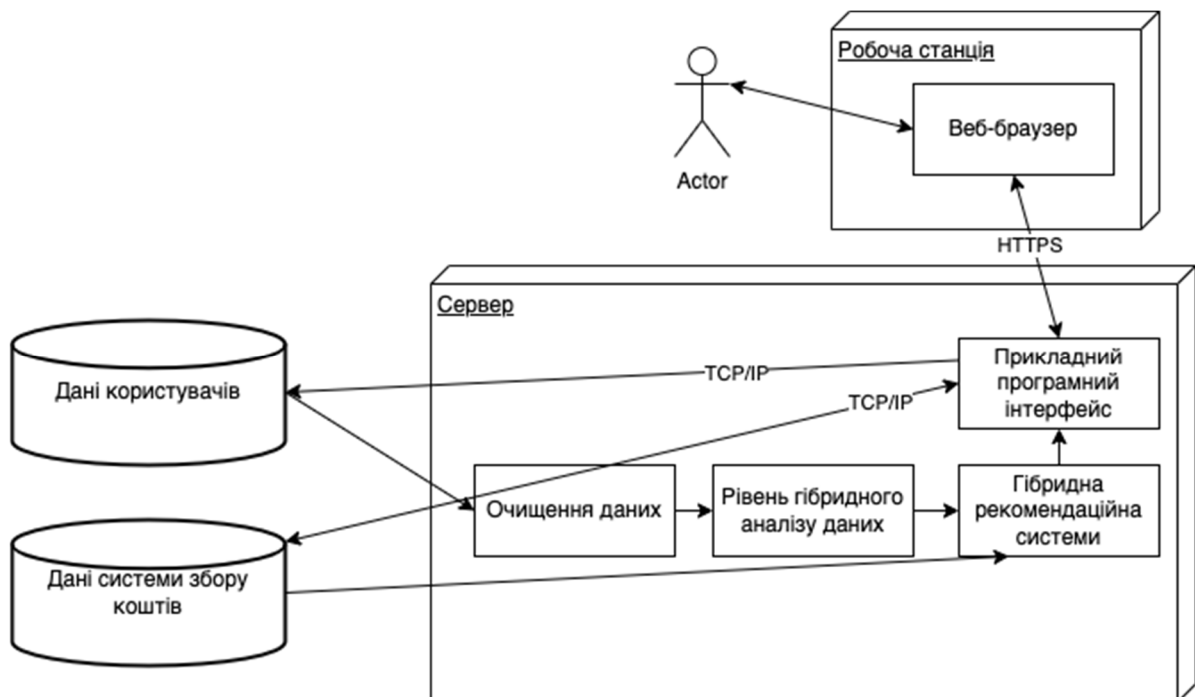


Рис. 1. Архітектура системи

У базі даних користувачів зберігаються профілі користувачів, історичні дані про взаємодію, важливі для створення персоналізованих рекомендацій. База даних проєктів містить інформацію про доступні проєкти, ураховуючи їхні атрибути та метадані.

Очищувач даних відповідає за попереднє опрацювання та очищення неопрацьованих даних. Він виконує такі завдання, як нормалізація тексту, вилучення дублікатів і перетворення даних для забезпечення їхньої якості.

Рівень аналізу даних є ядром системи, складеної з трьох основних компонентів. Компонент “тематичне моделювання” використовує прихований розподіл Діріхле, щоб виявити приховані тематичні структури в даних, що дає змогу глибше зрозуміти інтереси користувачів і теми вмісту. Компонент “аналіз тональності тексту” зосереджується на вилученні та аналізі настроїв, виражених у публікаціях у соціальних мережах, що допомагає системі розрізнати емоційний тон, ставлення та думки в даних. Компонент “аналіз соціальних мереж” використовується для виявлення впливових осіб і спільнот користувачів, підвищення актуальності та обізнаності про елементи системи.

Рівень гібридних рекомендацій використовує результати рівня аналізу гібридних даних, зокрема тематичне моделювання, семантичний аналіз і АСМ, для створення персоналізованих рекомендацій. Він поєднує інформацію із цих компонентів, щоб надати користувачам індивідуальні пропозиції.

API слугує мостом між системою та зовнішніми інтерфейсами. Він надсилає та отримує дані до та з бази даних користувачів і бази даних проєктів, забезпечуючи їхнє збереження і оновлення в реальному часі. Інтерфейс браузера представляє точку взаємодії між користувачами та системою рекомендацій. Користувачі вводять свої вподобання та отримують рекомендації через цей інтерфейс.

6. Характерні риси запропонованого рекомендаційного алгоритму

Для підвищення ефективності рекомендацій алгоритм ґрунтується на розробленні гібридних моделей, які поєднують кращі риси розглянутих методів інтелектуального аналізу даних: АТТ, ТМ та АСМ. Інтегрування цих підходів допомагає повніше зрозуміти почуття, інтереси і соціальні зв'язки користувачів, що сприяє надаванню точніших і персоналізованіших рекомендацій.

Відгуки користувачів і постійне оцінювання важливі для вдосконалення надання рекомендацій. Активний пошук відгуків користувачів за допомогою опитувань, рейтингів або явних механізмів зворотного зв'язку дає змогу системі збирати цінну інформацію про вподобання та рівень задоволеності користувачів. Введення цих відгуків у рекомендаційний алгоритм допомагає адаптувати рекомендації до потреб і вподобань окремих користувачів. Постійне оцінювання ефективності системи рекомендацій за допомогою показників і тестування користувачами сприяє ітераційним покращенням та гарантує, що вона залишиться ефективною та надійною.

Врахування моделей рекомендацій, які можна інтерпретувати та пояснювати, сприяє довірі та розумінню користувачів. Побудова моделей, які чітко пояснюють свої рекомендації, допомагає користувачам зрозуміти основні причини та фактори, які впливають на пропозиції, зроблені системою. Завдяки обґрунтуванню рекомендацій користувачі можуть приймати обґрунтованіші рішення та відчувати більший контроль над своїм вибором. Інтерпретовані моделі також підвищують прозорість і підзвітність, даючи користувачам змогу оцінювати правильність і потенційні упередження системи.

Крім того, введення аналізу контексту в формування рекомендацій додатково підвищує персоналізацію та релевантність рекомендацій. Враховуючи контекстуальні фактори, такі як місцезнаходження, час і поведінку користувача, система надає індивідуальні рекомендації, які відповідають конкретним обставинам і потребам окремих користувачів [20]. Наприклад, рекомендації місцевих благодійних організацій або кампаній, що потребують часу, можуть збільшити ймовірність залучення користувачів і підтримки.

Врахування згаданих характерних рис дає змогу надавати точніші, персоналізованіші та етично обгрунтованіші рекомендації у системі. Завдяки постійним дослідженням та інноваціям галузь аналізу даних продовжує розвиватися, надаючи людям можливість приймати обгрунтовані рішення та робити вагомий внесок у проекти, які їм близькі. Це сприятиме ефективнішому розподілу ресурсів на благодійні цілі та, зрештою, позитивним змінам у суспільстві.

Висновки

Комбінований алгоритм, що ґрунтується на збиранні та аналізі даних з соціальних мереж та поєднує кращі риси методів інтелектуального аналізу даних (АТТ, ТМ та АСМ), сприяє ефективній та персоналізованій роботі системи рекомендацій. У статті проаналізовано відомі рішення щодо добування даних із соціальних мереж та їх використання у системах рекомендацій.

У сфері рекомендаційних систем поєднано застосування методів вилучення даних із соціальних мереж, зокрема АТТ, ТМ і АСМ, відкриває багато можливостей для покращення взаємодії з користувачем і якості рекомендацій. АТТ дає змогу отримувати персоналізовані рекомендації, розшифровуючи настрої та вподобання користувачів, гарантуючи, що запропоновані елементи відповідають позитивним настроям і уникаючи тих, які пов'язані із негативним ставленням. Оскільки інтеграція даних про настрої в гібридні процеси рекомендацій істотно покращує їхню якість, громадські організації отримують розуміння сприйняття їхніх ідей спільнотами, що дозволяє їм налаштовувати стратегії та краще взаємодіяти зі своєю аудиторією.

Аналогічно ТМ є потужним інструментом для розуміння інтересів і проблем користувачів, надання індивідуальних рекомендацій на підставі тем, що домінують у даних соціальних мереж. Здатність узгодити рекомендації із тематичними вподобаннями користувачів покращує взаємодію та сприяє значущому внеску. ТМ також доводить свою цінність у відстеженні тенденцій, що розвиваються, і дискусіях, гарантуючи, що система рекомендацій залишається синхронізованою з динамічністю громадських інтересів і проблем.

З іншого боку, АСМ відіграє ключову роль у виявленні впливових осіб, які можуть просувати елементи системи та рекомендувати відповідні можливості своїм підписникам, враховуючи її переваги (значно вищу точність, запам'ятовуваність та оцінку F1) порівняно із колаборативним фільтруванням. Досліджуючи спільноти, АСМ визначає кластери користувачів зі спільними інтересами, прокладаючи шлях до індивідуальних рекомендацій, які резонують із конкретними спільнотами.

Підсумовуючи, інтеграція АТТ, ТМ і АСМ у системи рекомендацій розширює їхню здатність надавати персоналізовані та контекстозалежні пропозиції, а також надає громадським організаціям цінну інформацію. Ці методи вилучення даних дають змогу системам рекомендацій не лише відповідати очікуванням користувачів, але й адаптуватися до ландшафту соціальних медіа, що постійно змінюється, сприяючи більш значущому та привабливішому досвіду користувачів.

Список літератури

1. Purva Grover, Arpan Kumar Kar, Yogesh Dwivedi, The evolution of social media influence – A literature review and research agenda, *International Journal of Information Management Data Insights*. URL: <https://doi.org/10.1016/j.jjime.2022.100116>
2. M. D. Devika, C. Sunitha, Amal Ganesh, Sentiment Analysis: A Comparative Study on Different Approaches, *Procedia Computer Science*. URL: <https://doi.org/10.1016/j.procs.2016.05.124>
3. AlBadani, Barakat & Shi, Ronghua & Dong, Jian.. A Novel Machine Learning Approach for Sentiment Analysis on Twitter Incorporating the Universal Language Model Fine-Tuning and SVM. *Applied System Innovation*. URL: https://www.researchgate.net/publication/357853465_A_Novel_Machine_Learning_Approach_for_Sentiment_Analysis_on_Twitter_Incorporating_the_Universal_Language_Model_Fine-Tuning_and_SVM
4. Chetashri Bhadane, Hardi Dalal, Heenal Doshi, Sentiment Analysis: Measuring Opinions, *Procedia Computer Science*. URL: <https://doi.org/10.1016/j.procs.2015.03.159>

5. Aisopos, Fotis & Papadakis, George & Varvarigou, Theodora. Sentiment analysis of social media content using N-Gram graphs. URL: https://www.researchgate.net/publication/254002881_Sentiment_analysis_of_social_media_content_using_N-Gram_graphs
6. Qianwen Ariel Xu, Victor Chang, Chrisina Jayne, A systematic review of social media-based sentiment analysis: Emerging trends and challenges, *Decision Analytics Journal*. URL: <https://doi.org/10.1016/j.dajour.2022.100073>
7. Kherwa, Pooja & Bansal, Poonam. (2018). Topic Modeling: A Comprehensive Review. *ICST Transactions on Scalable Information Systems*. URL: https://www.researchgate.net/publication/334667298_Topic_Modeling_A_Comprehensive_Review
8. Evangelopoulos, Nicholas & Zhang, Xiaoni & Prybutok, V.R. Latent Semantic Analysis: Five Methodological Recommendations. *European Journal of Information Systems*. [URL: https://www.researchgate.net/publication/236133983_Latent_Semantic_Analysis_Five_Methodological_Recommendations
9. Choo, Jaegul & Lee, Changhyun & Reddy, Chandan & Park, Haesun. (2013). UTOPIAN: User-Driven Topic Modeling Based on Interactive Nonnegative Matrix Factorization. *IEEE transactions on visualization and computer graphics*. URL: https://www.researchgate.net/publication/256837226_UTOPIAN_User-Driven_Topic_Modeling_Based_on_Interactive_Nonnegative_Matrix_Factorization
10. Alghamdi, Rubayyi & Alfalqi, Khalid. (2015). A Survey of Topic Modeling in Text Mining. *International Journal of Advanced Computer Science and Applications*. URL: https://www.researchgate.net/publication/276327703_A_Survey_of_Topic_Modeling_in_Text_Mining
11. Marianela García Lozano, Jonah Schreiber, Joel Brynielsson, Tracking geographical locations using a geo-aware topic model for analyzing social media data. *Decision Support Systems*. URL: <https://doi.org/10.1016/j.dss.2017.05.006>
12. Charifa Laghrifat, Mohamed Essalih, A Set of Measures of Centrality by Level for Social Network Analysis, *Procedia Computer Science*. URL: <https://doi.org/10.1016/j.procs.2023.01.348>
13. Williams, Trenton & Shepherd, Dean. (2015). Mix Method Social Network Analysis: Combining Inductive Concept Development, Content Analysis, and Secondary Data for Quantitative Analysis. *Organizational Research Methods*. URL: https://www.researchgate.net/publication/283709926_Mix_Method_Social_Network_Analysis_Combining_Inductive_Concept_Development_Content_Analysis_and_Secondary_Data_for_Quantitative_Analysis
14. Fatma Altuntas, Serkan Altuntas, Turkey Dereli, Social network analysis of tourism data: A case study of quarantine decisions in COVID-19 pandemic, *International Journal of Information Management Data Insights*. URL: <https://doi.org/10.1016/j.ijime.2022.100108>
15. Novita Hanafiah, Alexander Kevin, Charles Sutanto, Fiona, Yulyani Arifin, Jaka Hartanto, Text Normalization Algorithm on Twitter in Complaint Category, *Procedia Computer Science*. URL: <https://doi.org/10.1016/j.procs.2017.10.004>
16. Deena Abul-Fottouh, Melodie Yunju Song, Anatoliy Gruzd, Examining algorithmic biases in YouTube's recommendations of vaccine videos, *International Journal of Medical Informatics*. URL: <https://doi.org/10.1016/j.ijmedinf.2020.104175>
17. Mehdi Elahi, Danial Khosh Kholgh, Mohammad Sina Kiarostami, Mourad Oussalah, Soroush Saghari, Hybrid recommendation by incorporating the sentiment of product reviews, *Information Sciences*. URL: <https://doi.org/10.1016/j.ins.2023.01.051>
18. Dixon Prem Daniel Rajendran, Rangaraja P. Sundarraj, Using topic models with browsing history in hybrid collaborative filtering recommender system: Experiments with user ratings, *International Journal of Information Management Data Insights*. URL: <https://doi.org/10.1016/j.ijime.2021.100027>
19. N Vedavathi, R Suhas Bharadwaj, Deep Flamingo Search and Reinforcement Learning Based Recommendation System for E-Learning Platform using Social Media, *Procedia Computer Science*. URL: <https://doi.org/10.1016/j.procs.2022.12.022>
20. Hossein A. Rahmani, Yashar Deldjoo, Tommaso di Noia, The role of context fusion on accuracy, beyond-accuracy, and fairness of point-of-interest recommendation systems, *Expert Systems with Applications*. URL: <https://doi.org/10.1016/j.eswa.2022.117700>

References

1. Purva Grover, Arpan Kumar Kar, Yogesh Dwivedi, The evolution of social media influence – A literature review and research agenda, *International Journal of Information Management Data Insights*. URL: <https://doi.org/10.1016/j.ijime.2022.100116>
2. M. D. Devika, C. Sunitha, Amal Ganesh, Sentiment Analysis: A Comparative Study on Different Approaches, *Procedia Computer Science*. URL: <https://doi.org/10.1016/j.procs.2016.05.124>

3. AlBadani, Barakat & Shi, Ronghua & Dong, Jian.. A Novel Machine Learning Approach for Sentiment Analysis on Twitter Incorporating the Universal Language Model Fine-Tuning and SVM. *Applied System Innovation*. URL: https://www.researchgate.net/publication/357853465_A_Novel_Machine_Learning_Approach_for_Sentiment_Analysis_on_Twitter_Incorporating_the_Universal_Language_Model_Fine-Tuning_and_SVM
4. Chetashri Bhadane, Hardi Dalal, Heenal Doshi, Sentiment Analysis: Measuring Opinions, *Procedia Computer Science*. URL: <https://doi.org/10.1016/j.procs.2015.03.159>
5. Aisopos, Fotis & Papadakis, George & Varvarigou, Theodora. Sentiment analysis of social media content using N-Gram graphs. URL: https://www.researchgate.net/publication/254002881_Sentiment_analysis_of_social_media_content_using_N-Gram_graphs
6. Qianwen Ariel Xu, Victor Chang, Chrisina Jayne, A systematic review of social media-based sentiment analysis: Emerging trends and challenges, *Decision Analytics Journal*. URL: <https://doi.org/10.1016/j.dajour.2022.100073>
7. Kherwa, Pooja & Bansal, Poonam. (2018). Topic Modeling: A Comprehensive Review. *ICST Transactions on Scalable Information Systems*. URL: https://www.researchgate.net/publication/334667298_Topic_Modeling_A_Comprehensive_Review
8. Evangelopoulos, Nicholas & Zhang, Xiaoni & Prybutok, V. R. Latent Semantic Analysis: Five Methodological Recommendations. *European Journal of Information Systems*. URL: https://www.researchgate.net/publication/236133983_Latent_Semantic_Analysis_Five_Methodological_Recommendations
9. Choo, Jaegul & Lee, Changhyun & Reddy, Chandan & Park, Haesun (2013). UTOPIAN: User-Driven Topic Modeling Based on Interactive Nonnegative Matrix Factorization. *IEEE transactions on visualization and computer graphics*. URL: https://www.researchgate.net/publication/256837226_UTOPIAN_User-Driven_Topic_Modeling_Based_on_Interactive_Nonnegative_Matrix_Factorization
10. Alghamdi, Rubayyi & Alfalqi, Khalid (2015). A Survey of Topic Modeling in Text Mining. *International Journal of Advanced Computer Science and Applications*. URL: https://www.researchgate.net/publication/276327703_A_Survey_of_Topic_Modeling_in_Text_Mining
11. Marianela García Lozano, Jonah Schreiber, Joel Brynielsson, Tracking geographical locations using a geo-aware topic model for analyzing social media data, *Decision Support Systems*. URL: <https://doi.org/10.1016/j.dss.2017.05.006>
12. Charifa Laghrifat, Mohamed Essalih, A Set of Measures of Centrality by Level for Social Network Analysis, *Procedia Computer Science*. URL: <https://doi.org/10.1016/j.procs.2023.01.348>
13. Williams, Trenton & Shepherd, Dean. (2015). Mix Method Social Network Analysis: *Combining Inductive Concept Development, Content Analysis, and Secondary Data for Quantitative Analysis*. *Organizational Research Methods*. URL: https://www.researchgate.net/publication/283709926_Mix_Method_Social_Network_Analysis_Combining_Inductive_Concept_Development_Content_Analysis_and_Secondary_Data_for_Quantitative_Analysis
14. Fatma Altuntas, Serkan Altuntas, Turkey Dereli, Social network analysis of tourism data: A case study of quarantine decisions in COVID-19 pandemic, *International Journal of Information Management Data Insights*. URL: <https://doi.org/10.1016/j.jjime.2022.100108>
15. Novita Hanafiah, Alexander Kevin, Charles Sutanto, Fiona, Yulyani Arifin, Jaka Hartanto, Text Normalization Algorithm on Twitter in Complaint Category, *Procedia Computer Science*. URL: <https://doi.org/10.1016/j.procs.2017.10.004>
16. Deena Abul-Fottouh, Melodie Yunju Song, Anatoliy Gruz, Examining algorithmic biases in YouTube's recommendations of vaccine videos, *International Journal of Medical Informatics*. URL: <https://doi.org/10.1016/j.ijmedinf.2020.104175>
17. Mehdi Elahi, Danial Khosh Kholgh, Mohammad Sina Kiarostami, Mourad Oussalah, Soroush Saghari, *Hybrid recommendation by incorporating the sentiment of product reviews*, *Information Sciences*. URL: <https://doi.org/10.1016/j.ins.2023.01.051>
18. Dixon Prem Daniel Rajendran, Rangaraja P Sundarraj, Using topic models with browsing history in hybrid collaborative filtering recommender system: Experiments with user ratings, *International Journal of Information Management Data Insights*. URL: <https://doi.org/10.1016/j.jjime.2021.100027>
19. N Vedavathi, R Suhas Bharadwaj, Deep Flamingo Search and Reinforcement Learning Based Recommendation System for E-Learning Platform using Social Media, *Procedia Computer Science*. URL: <https://doi.org/10.1016/j.procs.2022.12.022>

20. Hossein A. Rahmani, Yashar Deldjoo, Tommaso di Noia, The role of context fusion on accuracy, beyond-accuracy, and fairness of point-of-interest recommendation systems, *Expert Systems with Applications*. URL: <https://doi.org/10.1016/j.eswa.2022.117700>

**FEATURES OF RECOMMENDATION ALGORITHM
ON BASE OF ANALYSIS OF SOCIAL NETWORK DATA
MINING METHODS**

Petro Smachylo¹, Liubov Zhuravchak²

Lviv Polytechnic National University,
Software Engineering Department, 12, S. Bandery str., Lviv, Ukraine
¹ petro.i.smachylo@lpnu.ua, ORCID 0009-0008-1285-4181,
² liubov.m.zhuravchak@lpnu.ua, ORCID 0000-0002-1444-5882

© Smachylo P., Zhuravchak L., 2023

In recent years, social media platforms have become powerful data collection tools to improve user experience. The vast amount of data generated through social media provides a unique opportunity to develop innovative recommendation systems. This article analyzes the application of data mining methods for social networks in the context of effective recommendation systems, focusing on three key methodologies: sentiment analysis (SA), topic modeling (TM) and social network analysis (SNA), highlighting their positive features.

SA allows the system to tailor recommendations based on sentiment analysis, offering users items that match their expressed emotions. Experiments show significant improvements in recommendation accuracy when sentiment data is integrated. TM allows the system to understand the main concerns of users by identifying dominant themes, thereby providing personalized recommendations and staying abreast of evolving trends. At the same time, AFM identifies influential users and communities, increasing relevance and awareness of system elements.

The article highlights the enormous potential of social networks for the development of effective, personalized recommendation systems. Using sentiment analysis and topic modeling, these systems can provide personalized and relevant recommendations based on public sentiment, trending topics, and social network dynamics.

Keywords: data analysis; social networks; recommendation system; semantic analysis; topic modeling.