

ранньою працею про оптимальну нелінійну проблему фільтрування Руслана Стратоновича, який першим описав послідовно-зворотній алгоритм.

Основна ідея ПММ є прогнозування відстані між позиціями мови подальшого джерела. З іншого боку, IBM модель №4 намагається передбачити відстань між позиціями мови цільової мови.

Список використаних джерел:

1. Chen, Дж., Паралельний Web J-Y.: Обробка тексту для Cross-Language IR. У працях RIAO-2000: «Content-Based Multimedia доступу до інформації», Париж, 12-14 квітня 2000 року.
2. IBM alignments models – Wikipedia [Електронний ресурс], – Режим доступу: https://en.wikipedia.org/wiki/IBM_alignment_models.
3. Гельфейнбейн І.Г., Гончарук О.В., Лехельт В.П., Ліпатов А.А., Шило В.В. Автоматичний переклад семантичної мережі WORDNET на російську мову. // Праці конференції «Діалог». М.: 2003.

Рибак О.О.

студент,

Національний технічний університет України

«Київський політехнічний інститут імені Ігоря Сікорського»

АНАЛІЗ АКТИВНОЇ АРХІТЕКТУРИ КЕШУВАННЯ БЕЗДРОТОВОЇ МЕРЕЖІ

В цій статі буде представлено нову практичну архітектуру мобільної мережі, яка здатна збирати, аналізувати та швидко пристосовуватись до різних типів даних. Ця архітектура керується за допомогою поведінки людини. Запропонована система здійснює збір контекстної інформації (наприклад, перегляд історії інформації о місцезнаходженні) та прогнозує дії користувачів просторово-часовим запитом кеша вибраної системою на границі мережі. Запропонована архітектура мобільної мережі дозволить розділити обчислення та виконання прогнозування контенту на ядрі мережі та кешуванні на базовій станції. Це дозволить задовольнити потреби користувачів як у продуктивності, так і у високій якості [1].

На рис. 1 показана комбінована мережева архітектура, де платформа великих даних розгорнута в ядрі мережі і знаходиться у режимі відстеження та прогнозування запитів споживачів, з увімкненим механізмом кешування на BSS.

Наприклад, є невелика мережа, яка складається з n маленьких комірок, де транспортний зв'язок є бездротовим. Інформацію, яку запитує користувач залишається в бібліотеці. Зміст бібліотеки буде поступово заповнюватись запитом користувача для того, щоб зрозуміти, що в майбутньому користувач захоче дізнатися. Також саме такий спосіб допоможе розвантажити транспортні мережі малих базових станцій, які обладнані ємнісним кешем контенту з бібліотеки. Але із-за величезного об'єму контенту у користувачів, системі буде

дуже важко обробити корисний користувацький контент, тому будемо використовувати процедуру дзеркального відображення ініціалізації для передачі потокового сліду на платформу великих даних в основній мобільній мережі.

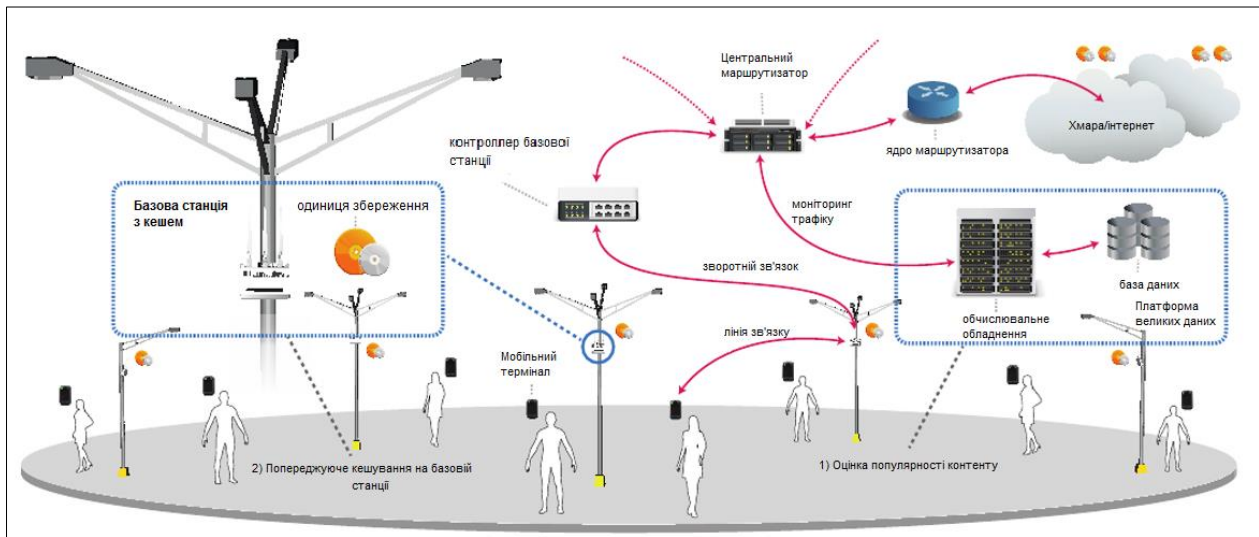


Рис. 1. Інформація переміщуються від хмари на BSS шляхом визначення стратегічного вмісту на платформі даних усередині ядра мережі, потім активно зберігає їх в кеші з підтримкою BSS

Можемо відмітити результати Турецьких дослідників, які показали, що у Туреччині з 10 регіональних напрямків основного мобільного оператора зв'язку у годину пік спостерігається швидкість передачі даних в 200 Мбіт/с, а загальний середній трафік по всіх регіонах складає понад 20 мільярдів пакетів щодня, а у вихідні дні 15 мільярдів пакетів [2].

На початку розмір сховища не буде задовольняти користувачів, тому що на даному етапі бібліотека має обмежений об'єм інформації. Експериментальні запити представлені на рисунку 2. Бібліотека має поки що максимальний об'єм розміром в 17.7 Гб. При порівнянні двох графіків при 0% розміру сховища буде відповідати 100% кешування всієї бібліотеки. Задоволеність користувачів буде зростати з об'ємом бібліотеки, також задоволеність досягається за допомогою повного накопичення популярної інформації в бібліотеці, яка буде фіксуватися параметрами системи, де система буде збирати статистику популярного контенту та зіставляти з запитом користувачів, щоб за допомогою такого методу досягти 100% задоволеності. На рисунку спостерігається розрив в продуктивності, який досягається до 79% від розміру сховища. Причина такого розриву полягає із-за помилки в оцінці контенту. Коли на базовій станції зайнято 40% від 17.7 Гб розміру сховища то кешування буде досягатися до 89%, а продуктивність буде залишатися на рівні 75-79% [3].

На рисунку 2 представлений метод, який дає набагато менше транспортного навантаження. Обидва методи розвантажують мережу на 98 відсотків. Якщо порівняти Ground Truth та Collaborative Filtering то Ground Truth буде мати перевагу над Collaborative Filtering так як цей метод має повну

інформацію о популярності контенту та може передбачити запити користувачів.

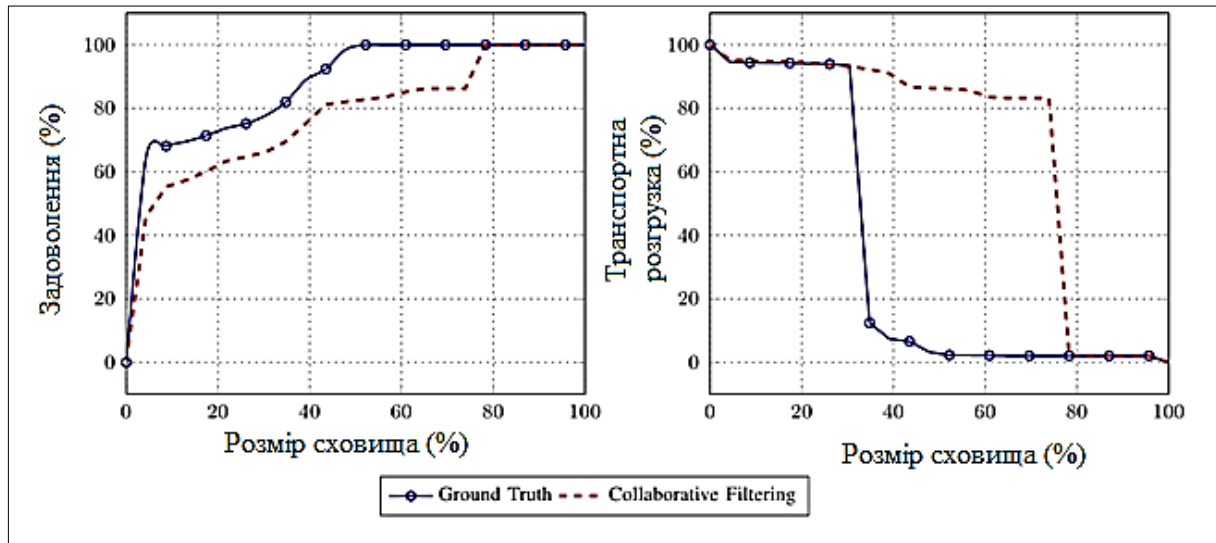


Рис. 2. Результати кешування на базових станціях

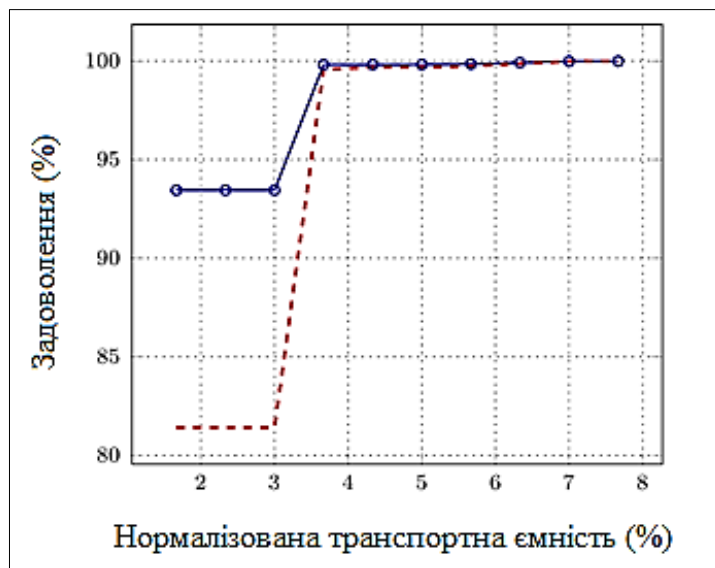


Рис. 3. Динаміка задоволеності користувачів за відношенням до нормалізованої транспортної ємності

Рисунок 3 ілюструє еволюцію задоволеності запитів споживачів к коефіцієнту потужності передачі, яка визначається, як відношення сумарної транспортної розгрузки ємності по відношенню до бездротової мережі [4].

Було проведено аналіз активної архітектури кешування бездротової мережі в результаті обробки величезної кількості даних на платформі великих даних. Завдяки добутої інформації даних трафіка будемо отримувати вигравш від кешування, але надалі буде продовжено дослідження. Одним з наступних можливих напрямів цієї роботи є вивчення великої бази аналізу даних в реальному часі.

Список використаних джерел:

1. G. Paschos, E. Bas̄etug̃, I. Land, G. Caire, and M. Debbah, «Wireless caching: Technical misconceptions and business barriers,» arXiv preprint arXiv:1602.00173, 2016.
2. X. Wang, M. Chen, T. Taleb, A. Ksentini, and V. Leung, «Cache in the Air: Exploiting content caching and delivery techniques for 5G systems,» IEEE Communications Magazine, vol. 52, no. 2, pp. 131–139, February 2014.
3. M. Tao, E. Chen, H. Zhou, and W. Yu, «Content-centric sparse multicast beamforming for cache-enabled cloud RAN,» [Online] arXiv: 1512.06938, 2015.
4. F. Bonomi, R. Milito, J. Zhu, and S. Addepalli, «Fog computing and its role in the internet of things».

Різник Р.К.

студент,

Харківський національний університет радіоелектроніки

РОЗРОБКА ТА ДОСЛІДЖЕННЯ МЕТОДІВ ВИЯВЛЕННЯ ПЛАГІАТУ В ТЕКСТОВИХ ДОКУМЕНТАХ

Швидкий розвиток інформаційних технологій надає змогу сучасним студентам та викладачам застосовувати метод написання робіт, що отримав назву «сору paste» – копіювання матеріалів з інтернету з їх мінімальним редагуванням. Копіювання наукових робіт без їх опрацювання – найбільш поширена форма наукової несумлінності. Дуже швидко йде руйнування наукової думки в академічній спільноті [1]. За час існування інформаційних технологій було запропоновано безліч методів для виявлення плагіату, включаючи методи засновані на простому порівнянні документів, аналізі термінів що зустрічаються в тексті, аналізі цитат, аналізі мовних стилів та інших. Часто для підвищення ефективності ці методи комбінують.

У даній роботі розглядаються підходи до класифікації показників схожості текстових документів а також розглядаються методи виявлення плагіату в текстових документах та їх комбінація [2].

Зазвичай системи виявлення плагіату базуються на порівнянні двох або більше документів. Для того, щоб порівняти документи і визначити ступінь їх схожості, необхідно кожному документу присвоїти відсоткове представлення унікальності. Існує декілька класифікацій метрик схожості документів. Одні з них базуються на загальній кількості документів, що залучені до аналізу, інші на обчислювальній складності методів визначення схожості. Метрики можна класифікувати за кількістю вимірів, що залежать від кількості документів. Також можна виділити поверхневі і структурні метрики. Поверхневою метрикою вимірюється схожість документів, простим порівняннями. Структурна метрика, навпаки, передбачає аналіз і опрацювання мовних особливостей. Класифікація показників схожості документів базується на семантичних та статистичних методах.