

ОПТИМІЗАЦІЯ ЗАДАЧІ РОЗСИЛАННЯ ЕЛЕКТРОННОЇ КОРЕСПОНДЕНЦІЇ

Поворозник В.С.

Тернопільський національний економічний університет, магістрант

I. Постановка проблеми

З задачею розсилання електронної кореспонденції люди стикаються дуже часто, для популяризації свого бізнесу, для автоматизації робочого процесу і т.д. Оскільки розсилання електронної кореспонденції не є безкоштовним процесом, то постає питання в ефективному розподіленні отримувачів цієї кореспонденції, для того щоб вона була надіслана тільки тим кому це насправді потрібно.

Вирішити дану задачу допоможе одна із задач інтелектуального аналізу даних (Data Mining) – кластеризація.

II. Мета роботи

Мета роботи полягає у створенні програмного забезпечення, яке на основі застосування одного із методів кластерного аналізу розподілятиме користувачів та електронну кореспонденцію на групи і в залежності від розподілу груп буде надсилати повідомлення.

III. Реалізація завдання

Для реалізації завдання обрано алгоритм ієрархічної кластеризації, знизу-вверх (агломеративний алгоритм), а саме алгоритм Ланса-Вільямса.

Спочатку кожен об'єкт вважається окремим кластером. Для кластерів визначається функція відстані:

$$R(\{x\}, \{x'\}) = p(x, x') \quad (1)$$

Потім запускається процес злиття. На кожній ітерації замість пари самих близьких кластерів U і V утворюється новий кластер $W = U \cup V$. Відстань від нового кластера W до будь-якого іншого кластера S обчислюється по відстанях $R(U, V)$, $R(U, S)$ і $R(V, S)$, які до цього моменту вже повинні бути відомі:

$$R(U \cup V, S) = \alpha_U R(U, S) + \alpha_V R(V, S) + \beta R(U, V) + \gamma |R(U, S) - R(V, S)| \quad (2)$$

Загальний алгоритм виглядає так:

1: Ініціалізація кластерів C_1 :

$$t := 1; C_t = \{\{x_1, \dots, x_\ell\}\}; \quad (3)$$

2: для всіх $t = 2, \dots, \ell$ (t -номер ітерації)

3: знайти в C_{t-1} два найближчих кластера:

$$(U, V) := \arg \min_{U \neq V} R(U, V); \quad (4)$$

$$R_t := R(U, V); \quad (5)$$

4: вилучити кластери U і V , додати утворений новий кластер $W = U \cup V$:

$$C_t := C_{t-1} \cup \{W\} \setminus \{U, V\}; \quad (6)$$

5: для всіх $S \in C_t$

6: вчислити відстань $R(W, S)$ по формулі Ланса-Вільямса

Для обчислення відстані обрано формулу відстані дальнього сусіда:

$$R(W, S) = \max_{w \in W, s \in S} p(w, s); \quad \alpha_U = \alpha_V = \frac{1}{2}, \beta = 0, \gamma = \frac{1}{2}. \quad (7)$$

Об'єктами виступають ключові слова, що містяться у даних отримувачів, і у самій кореспонденції. Спочатку кластеризуються дані кореспонденції, після чого дані отримувачів, далі відбувається злиття і виділення спільних кластерів, тоді відповідно до кластерів відбувається надсилання кореспонденції. У системі будуть передбачені методи для перегляду/редагування даних

отримувачів, і надісланих об'єктів кореспонденції, буде можливість завантаження даних з файлів. Також будуть зберігатись усі дані про надсилання.

Висновок

Запропоновано застосувати алгоритм ієрархічної кластеризації для оптимізації задачі розсилання електронної кореспонденції, а саме розбиття отримувачів і кореспонденцію на групи (кластери), знаходження спільних груп і в результаті розсилання кореспонденції тільки тим хто найбільш зацікавлений у ній.

Ця система дозволить краще оптимізувати процес розповсюдження інформації в колах користувача системи.

Список використаних джерел

1. Лекции по алгоритмам кластеризации и многомерного шкалирования [Електронний ресурс] //К. В. Воронцов - 21.12.2007. - 18 с. Режим доступу до ресурсу: <http://www.ccas.ru/voron/download/Clustering.pdf>.
2. Обзор алгоритмов кластеризации данных [Електронний ресурс] – 2010 - Режим доступу до ресурсу: <https://habrahabr.ru/post/101338/>
3. Применение методов кластеризации для обработки новостного потока [Електронний ресурс] – Режим доступу до ресурсу: <http://www.moluch.ru/conf/tech/archive/2/207/>

УДК 004.62

ПІДВИЩЕННЯ ЕФЕКТИВНОСТІ ПРОЦЕСУ ЗАПОВНЕННЯ СХОВИЩА ДАНИХ ДЛЯ СИСТЕМИ ТАЙМ МЕНЕДЖМЕНТУ

Пойдич В.С.¹⁾, Струбицька І.П.²⁾

Тернопільський національний економічний університет

¹⁾ магістрант; ²⁾ к.т.н., доцент

I. Постановка проблеми

Тайм менеджмент – один з найважливіших критеріїв успішності у бізнесі. «Час це гроші», - не даремно так кажуть. Тому в період новітніх технологій необхідно максимально автоматизувати процеси, які більшою мірою є марудними при ручному виконанні, але які є істотними для успішності в різних сферах діяльності. Розподіл часу використовується всюди: від простого нагадування про те що і коли треба зробити, до потужного графіку, від якого залежить успішність підприємства. І як вже було сказано, заповнення сховища даних іноді є довгим процесом, що безпосередньо потребує ще часу, який би міг бути використаний на інші важливіші справи. Було розроблено багато методологій по правильному менеджменту часу, як і коли отримувати найкращий результат, як виділяти категорії, як все ж таки встигати і керувати бізнесом і проводити час з сім'єю [1-3].

II. Мета роботи

Метою дослідження є розробка алгоритму, який буде виконувати функцію ефективного розподілу часу відносно певних критеріїв, які можуть бути задані користувачем або визначені системою, і відповідно до результатів будувати максимально гнучкий графік виконання поставлених задач. Також система буде сама навчатись відносно реакцій користувача.

III. Особливості програмної реалізації продукту

Система буде реалізована на мові C# з використанням технології Xamarin [4], оскільки вона буде орієнтована на мобільні платформи з ОС Android або iOS. Програмна система буде розроблена з використанням алгоритмів, які будуть активно слухати і запам'ятовувати вибори користувача, відносно однієї або іншої категорії. При розробці, як основний критерій, було вирішено використати вже готове розділення Ейзенхауера. Логіка полягає в тому що завдання діляться на 4 категорії:

1. Термінові-важливі;
2. Термінові-неважливі;
3. Нетермінові-важливі;
4. Нетермінові – неважливі.

Саме такий розподіл дозволяє найкраще відобразити все сховище даних, і краще аналізувати дані та подавати користувачу найвідповідніші задачі. На рис. 1 показано куб Ейзенхауера.