

УДК 004.89

ЗМЕНШЕННЯ РОЗМІРНОСТІ МАТРИЦІ "КОРИСТУВАЧ-ОБ'ЄКТ" ПРИ КОЛАБОРАТИВНІЙ ФІЛЬТРАЦІЇ

Савчук Тамара, Сакалюк Антон

Вінницький національний технічний університет, Україна

Анотація

Основним недоліком алгоритмів колаборативної фільтрації є необхідність виконання великої кількості операцій для обчислення ступеня схожості об'єктів або користувачів при прогнозуванні невідомого рейтингу. Для зменшення трудомісткості операцій при обчисленні ступеня схожості використовується підхід пониження розмірності матриці користувач-об'єкт, заснований на розкладанні цієї матриці по сингулярним значенням.

Abstract

The main disadvantage of collaborative filtering algorithms is the need to perform a large number of operations to calculate the degree of similarity of objects or users in predicting unknown rating. To reduce the complexity of operations when calculating the degree of similarity used dimension reduction matrix user-object based on the decomposition of this matrix by singular value.

Рекомендаційні системи застосовують методи виявлення знань до проблеми створення рекомендації певних об'єктів (фільми, відео, книги, зображення, веб-сторінки) для користувачів.

Колаборативна фільтрація використовує схожість думок різних користувачів для видачі рекомендацій щодо об'єктів. Рекомендаційні системи на основі колаборативної фільтрації - зручна альтернатива класичним пошуковим алгоритмам, так як використовують фактори, які неможливо отримати з технічного аналізу інформації. Тому впровадження подібних механізмів прискорить процес пошуку потрібної інформації, збільшить її повноту і достовірність.

Для зменшення складності обчислення ступеня схожості векторів об'єктів або користувачів використовується підхід пониження розмірності матриці "користувач-об'єкт" заснований на розкладанні цієї матриці по сингулярним значенням. Розкладання по сингулярним значенням (SVD - Singular Value Decomposition) є подання матриці $R \in Mat(n, m)$ у вигляді

$$R = U \times S \times V^T \quad (1)$$

де матриці $U \in Mat(n, r)$ та $V \in Mat(m, r)$ складаються з ортонормальних стовпців, які є власними векторами при ненульових власних значеннях матриць RR^T і $R^T R$ відповідно, а

$$S = \left\{ \begin{array}{cccc} \lambda_1 & 0 & \dots & 0 \\ 0 & \lambda_2 & \dots & 0 \\ \dots & \dots & \dots & \dots \\ 0 & 0 & \dots & \lambda_r \end{array} \right\} \in Mat(r, r) \quad (2)$$

- діагональна матриця з додатними діагональними елементами $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_r > 0$, відсортованими в порядку спадання. Замість n -мірного вектора j -ого об'єкту Y_j ми розглядаємо d -мірний вектор, що представляє собою вектор коефіцієнтів розкладання проекції Y_j по базису U_d . Використовуючи описаний підхід, для визначення ступеня схожості векторів об'єктів Y_i і Y_j ми обчислюємо ступінь схожості їх d -мірних апроксимацій.

На відміну від традиційного підходу обчислення схожості всіх об'єктів, нам потрібно $O(d)$ операцій для обчислення ступеня схожості між апроксимаціями векторів об'єктів замість $O(n)$, що значно прискорює обчислення при $d \ll n$.

Загальний алгоритм зменшення розмірності при колаборативній фільтрації виглядатиме:

1. Знизити розрідженість вихідної матриці, заповнюючи її середніми оцінками об'єктів.
2. Нормалізувати кожен запис в матриці середніми оцінками користувача.
3. Застосувати зменшення розмірності SVD:

$$R = U \times S \times V^T \rightarrow R_d = U_d \times S_d \times V_d^T \quad (3)$$

SVD підвищує масштабованість рекомендаційної системи. Зменшення розмірності дозволяє підвищити швидкість обробки даних, оскільки всі обчислення виконуються на підготовчому етапі, тобто не при запиті користувача, а заздалегідь.

Список використаних джерел:

1. Su, Xiaoyuan. A Survey of Collaborative Filtering Techniques / Xiaoyuan Su, Taghi, M. Khoshgoftaar. - Hindawi Publishing Corporation USA, 2009. - 215 p.
2. Т.О. Савчук, А.В. Сакалюк. Застосування кластерного аналізу для вдосконалення алгоритму колаборативної фільтрації. / Вісник Хмельницького національного університету, 2011. №1.
3. Т. Сегаран. Программируем коллективный разум; пер. с англ. А. Слинкина - СПб: Символ-Плюс, 2008. - 368 с.