

ANALISIS DAN IMPLEMENTASI CLASS-BASED COLLABORATIVE FILTERING PADA RECOMMENDER SYSTEM

Aditya Hermawan¹, Angelina Prima Kurniati², Kemas Rahmat Saleh Wiharja³

¹Teknik Informatika, Fakultas Teknik Informatika, Universitas Telkom

Abstrak

Recommender system adalah sistem yang dapat digunakan untuk memprediksi sebuah items dalam hal ini berupa movies, berdasarkan informasi yang diperoleh dari user, sehingga didapatkan rekomendasi berdasarkan profil penggunaannya. Collaborative filtering merupakan pendekatan pada recommender system yang merekomendasikan items dengan mencari similarity antara user atau antara item berdasarkan informasi yang sudah ada pada user atau item lainnya.

Tugas Akhir ini mengimplementasikan dan menganalisis performansi classbased collaborative filtering pada recommender system. Algoritma class-based merupakan pengembangan dari user-based collaborative filtering. Algoritma class-based dalam memprediksi nilai rating suatu item dengan menggabungkan dua konsep yaitu matrix user-class dan instance selection. Sehingga hasil prediksi items yang akan didapatkan menjadi maksimal. Data yang digunakan adalah data set movielens. Faktor-faktor yang digunakan dalam analisis adalah user frequency threshold, given k, dan Ncommon. Tugas akhir ini menganalisis tingkat akurasi prediksi rating yang dihasilkan dengan metoda evaluasi MAE (Mean Absolut Error).

Faktor-faktor seperti user frequency threshold, given k, dan Ncommon mempengaruhi tingkat akurasi prediksi rating dinilai dari MAE-nya. Penerapan user frequency threshold dan given k pada algoritma class-based akan cenderung berdampak lebih baik MAE-nya daripada yang tidak diterapkan didalamnya.

Kata Kunci : recommender system, collaborative filtering, class-based, instance

Abstract

Recommender system is a system that can be used to predict a items in this case in the form of movies, based on the information obtained from the user, so that the recommendations based on the user's profile was obtained. Collaborative filtering in recommender system approach is the recommended items by finding similarity between users or between items on the basis of the information already exists on the user or other items.

This final task implements and analyzes the performance of class-based collaborative filtering in recommender system. Class-based algorithms is the development of user-based collaborative filtering. Class-based algorithms to predict the rating an item value by combining two concepts, namely matrix userclass and instance selection. So the prediction results obtained shall become items maximum. Data used is data set movielens. Factors that are used in the analysis are the user frequency threshold, given k and Ncommon. The final task is to analyze the level of accuracy prediction rating produced by the method of evaluation of MAE (Mean Absolute Error).

Factors such as user frequency threshold, the given k, and Ncommon affecting the level of accuracy prediction rating based its MAE. The application of user frequency threshold and the given k on algorithms class-based tend impact better of the MAE than not applied it.

Keywords : recommender system, collaborative filtering, class-based, instance

BAB I

Pendahuluan

1.1 Latar Belakang

Bagi sebagian orang, setelah mereka menemukan suatu hal yang sesuai dengan kepribadiannya dan memberikan penilaiannya terhadap hal itu lalu setelah itu ingin mendapatkan rekomendasi hal serupa yang belum mereka ketahui tetapi bisa sesuai dengan kepribadian mereka maka itu jadi satu masalah tersendiri. Oleh sebab itu, solusi yang diberikan adalah suatu *recommender system*. *Recommender system* adalah sistem yang dapat digunakan untuk memprediksi sebuah *items* dalam hal ini berupa *movies*, berdasarkan informasi yang diperoleh dari user, sehingga didapatkan rekomendasi berdasarkan profil penggunaanya.

Ada tiga kategori pendekatan dalam *recommender system* yaitu *rule-based systems*, *content-based filtering systems*, dan *collaborative filtering systems*. Pendekatan *recommender system* dengan *content-based* memiliki karakteristik yaitu mencari item berdasarkan konten *items* yang telah di-rate sebelumnya oleh user [3]. Oleh karena itu memiliki keterbatasan ketika user meminta rekomendasi item yang berbeda kontennya dengan *items* yang pernah dipilih user tersebut. Sedangkan pendekatan *collaborative filtering* merekomendasikan item berdasarkan kemiripan satu user dengan user lainnya, misal pemberian *rating* pada suatu item. Oleh karena itu ada satu poin yang menjadikan *collaborative filtering* memiliki kelebihan dibanding *content-based* yaitu *collaborative filtering* dapat merekomendasikan *items* yang berbeda konten dengan item yang telah di-rate oleh user.

Pendekatan *collaborative filtering* merupakan salah satu pendekatan yang paling umum digunakan dalam *recommender system* [10]. *Collaborative Filtering* adalah proses mencari dan merekomendasikan item kepada *active user* berupa item yang belum pernah di-rating olehnya tetapi telah di-rating sangat bagus oleh user lainnya berdasarkan kesamaan selera pada item lainnya yang sama-sama telah mereka berikan *rating* [5]. Bagaimanapun, pendekatan tersebut masih mengalami dua masalah dasar yaitu *sparsity* dan *scalability* [10]. Permasalahan pada *sparsity* terjadi karena terlalu banyak item yang belum di-rating oleh user, akibatnya *matrix user-item* menjadi sangat renggang sehingga yang berdampak hasil prediksi kurang akurat. Sedangkan permasalahan pada *scalability* adalah kondisi dimana algoritma *k nearest neighbor* yang biasa dipakai pada klasik *collaborative filtering* memiliki keterbatasan mencari user *neighbors* untuk item *target* ketika data user dan item dalam *recommender system* akan terus bertambah.

Ada dua pendekatan dalam *collaborative filtering* yaitu *user-based* dan *item-based collaborative filtering*. Pada *user-based* merekomendasikan item-nya dengan mencari kemiripan selera seorang user dengan user lainnya, misal cara pemberian *rating* terhadap suatu item. Sedangkan pada *item-based* merekomendasikan item-nya dengan melihat kedekatan *target item* terhadap item-item yang telah di-rating oleh user. *User-based* adalah ide dasar dalam *collaborative filtering* yang telah sukses diterapkan dalam penelitian dan aplikasi seperti *information filtering* [10].

Metode *class-based collaborative filtering* merupakan pengembangan dari pendekatan *user-based collaborative filtering* yang diharapkan dapat meningkatkan performansi prediksi dan rekomendasi. Pengembangannya berupa penggunaan dua konsep matriks *user-class* dan *instance selection* guna mengatasi masalah *scalability*.

1.2 Perumusan Masalah

Dalam pengerjaan Tugas Akhir ini, ada beberapa rumusan masalah yang ingin diselesaikan, yaitu:

Bagaimana menerapkan metode *Class-Based Collaborative Filtering* pada *Recommender System*.

1. Bagaimana mengevaluasi hasil prediksi rating recommender system dengan metode *class-based collaborative filtering*.
2. Apa faktor-faktor yang mempengaruhi hasil prediksi pada *class-based CF*.

Batasan Masalah

Dalam implementasi tugas akhir ini dibatasi oleh beberapa hal sebagai berikut:

1. Dataset film yang dipakai adalah data MovieLens yang disediakan oleh *Grouplens Research Center*
2. *Rating* yang digunakan hanya numeric yaitu 0-5
3. *Item* yang dijadikan objek rekomendasi adalah film yang telah di-rating oleh user

1.3 Tujuan

Tujuan yang ingin dicapai dalam pengerjaan Tugas Akhir ini adalah :

1. Mengimplementasikan *class-based collaborative filtering* pada *recommender system*.
2. Mengukur dan menganalisis tingkat error hasil prediksi terhadap rating aslinya dari user dengan menggunakan *Mean Absolute Error* (MAE).
3. Menganalisis pengaruh *user frequency threshold*, protokol *given k*, dan *n common* pada *class-based collaborative filtering* terhadap ketepatan prediksi rating yang dihasilkan.

1.4 Metodologi penyelesaian masalah

1. Studi Literatur :
Pencarian referensi dan sumber-sumber yang berhubungan dengan *Recommender System*, *Collaborative Filtering* dan literatur-literatur yang berkaitan dengan *Class-Based*.
2. Pengumpulan data :
Mengumpulkan data-data *item* yang diperlukan sebagai *training set* dan *test set*.
3. Perancangan sistem:
Perancangan aplikasi yang akan dibangun, mencari metode yang akan digunakan untuk menyelesaikan permasalahan, termasuk menentukan bahasa pemrograman yang dipakai, arsitektur, fungsionalitas, dan antarmuka sistem
4. Implementasi dan Pembangunan Sistem
Tahap pembangunan *Recommender System* dengan melakukan pembangunan terhadap database, pembuatan algoritma, kemudian membuat *interface* aplikasi untuk memudahkan *user* dalam mengakses sistem.
5. Testing & Analisis
Melakukan analisis dari implementasi sistem dan pengujian hasil. *Mean Absolute Error* (MAE) digunakan untuk menghitung akurasi dari hasil prediksi rating
6. Penyusunan laporan Tugas Akhir
Menyusun laporan tertulis dan memberikan kesimpulan dari hasil penelitian yang dilakukan, pemberian saran untuk pengembangan sistem yang dibangun ke depannya.

BAB V

KESIMPULAN DAN SARAN

5.1 Kesimpulan

Berdasarkan hasil uji coba dan analisis yang telah dilakukan, maka dapat diambil kesimpulan sebagai berikut :

1. *User frequency threshold* berpengaruh dalam memperbaiki akurasi hasil prediksi. Hal tersebut dapat dilihat dari nilai MAE yang lebih baik (nilainya lebih kecil dibanding tanpa threshold).
2. Protokol *given k* mempengaruhi nilai MAE. Kecenderungannya semakin besar nilai *k* pada *given k* akan memperkecil nilai MAE. Ini berarti dengan nilai *given k* yang semakin besar, nilai rata-rata akurasi prediksi ratingnya akan semakin baik.
3. Variabel *n common* mempengaruhi nilai MAE dilihat dari pengujiannya yang menunjukkan semakin besar nilai *n common*, kecenderungannya nilai MAE semakin kecil. Semakin kecil nilai MAE artinya semakin baik akurasi prediksi *rating*-nya.

5.2 Saran

Setelah tugas akhir ini selesai dikerjakan dan telah dilakukan analisis, penulis memiliki beberapa saran sebagai berikut :

1. Oleh karena algoritma class-based masih bisa dikembangkan lagi maka terbuka kemungkinan penambahan algoritma lainnya yang dapat memprediksi rating secara cepat dan akurat dari dataset yang besar.
2. Pengujian terhadap class-based yang disajikan di tugas akhir ini hanya sedikit dari yang bisa dilakukan. Oleh sebab itu, peluang untuk pengujian dari sisi yang lainnya masih sangat mungkin dilakukan.

Telkom
University

Referensi

- [1] Breese, J. S., D. Heckerman and C. Kadie. "Empirical Analysis of Predictive Algorithms for Collaborative Filtering." In Proceedings of The 14th Conference on Uncertainty in Artificial Intelligence. 1998. 43-52.
- [2] Grouplens. 6 December 2011 <<http://www.grouplens.org/node/73>>.
- [3] Hofmann, T. and J. Puzicha. "Latent Class Models for Collaborative Filtering." In Proceedings of The 16th International Joint Conference on Artificial Intelligence. 1999. 688-693.
- [4] Li, Qing and Byeong Man Kim. "An Approach for Combining Content Based and Collaborative Filters." In Proceedings of The 6th International Workshop on Information Retrieval with Asian Languages-Vol.11. 2003. 17-24.
- [5] Pazzani, J. "A Framework For Collaborative, Content-Based, and Demographic Filtering." Book of Artificial Intelligence Review Vol. 13, Numbers 5-6. 1999. 393-408.
- [6] Teorey, T., S. Lightstone and T. Nadeau. Database Modeling and Design: Logical Design, 4th edition. Morgan Kaufmann press, 2005.
- [7] Ullman, J. and J. Widom. First Course in Database Systems. Prentice Hall Inc., 1997. 1.
- [8] Weber, R., H.J. Schek and S. Blott. "A Quantitative Analysis and Performance Study for Similarity-Search Methods in High-Dimensional Spaces." In Proceedings of The International Conference on Very Large Databases. 1998. 194-205.
- [9] Yang, Y and J.O. Pedersen. "A Comparative Study on Feature Selection in Text Categorization." In Proceeding of The 14th International Conference on Machine Learning. 1997. 412-420.
- [10] Yu, K., et al. "Feature Weighting and Instance Selection for Collaborative Filtering." In Proceedings of The 2nd International Workshop on Management of Information on The Web-Web Data and Text Mining. 2001.
- [11] Yu, K., X. Xu and J. Tao. "Instance Selection Techniques for Memory-Based Collaborative Filtering." In Proceedings of The 2nd SIAM International Conference on Data Mining. 2002. 59-74.
- [12] Zeng, C., Chun-Xiao, Xing and Li-Zhu Zhou. "Similarity Measure and Instance Selection for Collaborative Filtering." In Proceedings of The 12th International Conference on World Wide Web. 2003. 652-658.

Telkom
University