

PENGEMBANGAN COLLABORATIVE FILTERING DENGAN MENGGUNAKAN DEMOGRAPHIC DATA PADA RECOMMENDER SYSTEM

Imam Rohendi¹, Kusuma Ayu Laksitowening², Kemas Rahmat Saleh Wiharja³

¹Teknik Informatika, Fakultas Teknik Informatika, Universitas Telkom

Abstrak

Recommender system merupakan sebuah sistem yang dapat digunakan untuk memprediksi sebuah item berdasarkan informasi yang diperoleh dari user, sehingga didapatkan rekomendasi berdasarkan profil penggunaannya. Collaborative filtering merupakan teknik yang umum digunakan dalam recommender system akan tetapi hasil prediksi yang didapatkan masih kurang memuaskan karena dalam perhitungannya, tidak memperhatikan latar belakang / demografi user atau item. Oleh karena itu, diperlukan informasi tambahan agar hasilnya lebih baik, yaitu demographic data .

0Tugas akhir ini akan mengimplementasikan collaborative filtering recommender system yang digabungkan dengan demographic data sebagai solusi untuk memproses hasil prediksi rating yang lebih baik. Selain itu, tugas akhir ini juga menganalisis akurasi prediksi rating yang dihasilkan oleh recommender system dengan metoda evaluasi MAE (Mean Absolute Error). Parameter yang digunakan dalam analisis adalah parameter α , β dan γ . Data yang digunakan adalah data set IMDB.

Akurasi prediksi yang dihasilkan oleh algoritma collaborative filtering dengan demographic data lebih besar dibandingkan dengan collaborative filtering biasa. Performansi terbaik terjadi dimana jumlah neighborhood saat jumlah terbanyak pada kecenderungan user dalam merating items. Hasil rekomendasi pada algoritma demographic collaborative filtering pada recommender system menunjukkan kesesuaian antara genre items hasil rekomendasi dengan genre item yang telah diberi rating oleh active user.

Kata Kunci : recommender system, collaborative filtering, rating, demographic data, genre, MAE.

Abstract

Recommender system is a system that can be used to predict the items based on information obtained from users, so we get recommendations based on user profiles. Collaborative filtering is a technique commonly used in recommender system, but prediction results obtained is still less satisfactory because in its calculations, did not pay attention to the background / demographics of users or items. Therefore, it needs an additional information to better results, namely Demographic data.

This final project will implement a collaborative filtering recommender system in combination with Demographic Data as a solution to process the prediction results which is better ratings. In addition, this final project also analyzed the prediction accuracy of the ratings generated by the recommender system with the method of evaluation MAE (Mean Absolute Error). The parameters used in the analysis are α , β and γ parameters. The data used are IMDB data sets.

Prediction accuracy generated by collaborative filtering algorithms with Demographic data is larger than plain collaborative filtering. The best performance occurs where the number of neighborhood at the greatest number of trend of user to rate an item. Results of recommendations on Demographic collaborative filtering algorithms in the recommender system shows the conformity between the item genre from result of recommendation with item genre that has been rated by active user.

Keywords : recommender system, collaborative filtering, rating, demographic data, genre, MAE.

BAB I

Pendahuluan

1.1 Latar belakang

Pencarian informasi melalui internet sering dihadapkan dengan fenomena meningkatnya jumlah informasi atau sering disebut dengan *information overload*. Hal ini menjadikan masalah tersendiri terhadap pencari informasi atau *user*. Oleh karena itu, dibuatlah sistem untuk membantu *user* agar pencariannya lebih mudah serta sesuai dengan keinginan *user* tersebut. Sistem tersebut dinamakan *recommender system*.

Recommender system adalah suatu teknologi yang didesain untuk mempermudah *user* dalam menemukan suatu data yang mungkin sesuai dengan profil *user* secara cepat dan dapat mengurangi jumlah informasi yang terlalu banyak (*information overload*)[11]. Contoh studi kasus yang akan diambil dalam Tugas akhir ini adalah prediksi *rating* terhadap suatu film untuk memudahkan *user* memilih film tersebut.

Didalam *recommender system* sendiri terdapat berbagai macam metode untuk mendapatkan informasi yang sesuai dengan kebutuhan *user*. Salah satu metode yang paling berhasil dan banyak digunakan saat ini adalah *collaborative filtering*[10]. *Collaborative filtering* adalah proses penyaringan data dengan cara mengidentifikasi *user* yang memiliki kesamaan dengan yang diberikan kepada *user* lain pada masa lampau untuk menghasilkan prediksi *rating* film yang sesuai dengan keinginannya[11]. Tetapi, kelemahan dari *collaborative filtering* adalah identifikasi yang dilakukan hanya memperhatikan *rating* terdahulunya saja atau tidak memperhatikan latar belakang/demografi dari *user* atau film tersebut, sehingga prediksi yang dihasilkan masih kurang sesuai dengan keinginan *user*[12]. Oleh karena itu, cara yang tepat untuk memperbaiki kelemahan tersebut adalah dengan menambahkan informasi di dalam *demographic data*. *Demographic data* merupakan sumber informasi yang bertujuan untuk mengkategorikan *user* atau *item* berdasarkan atribut pribadinya dan membuat rekomendasi berdasarkan kelas rekomendasi[5].

Pada penyusunan tugas akhir ini akan dilakukan implementasi *collaborative filtering* yang digabungkan dengan *demographic data* sebagai solusi untuk memproses hasil prediksi *rating* suatu film yang lebih baik. Penggabungan antara kedua metode tersebut dinamakan *hybrid filtering*.

Salah satu analisis yang akan dilakukan adalah analisis ketepatan hasil rekomendasi yang dihasilkan oleh metode tersebut berdasarkan flag pada salah satu perhitungan *demographic data* (*enhanced correlation*) atau parameter α , β dan γ .

1.2 Perumusan masalah

Berdasarkan latar belakang yang telah dijelaskan sebelumnya, maka permasalahan yang akan dijabarkan dan diteliti antara lain:

1. Bagaimana membangun *recommender system collaborative filtering* yang sesuai dengan demografi *user* atau *item*.
2. Apakah *demographic collaborative filtering* menghasilkan prediksi yang lebih baik dari *collaborative filtering* biasa.
3. Bagaimana pengaruh parameter α , β dan γ pada perhitungan *enhanced correlation* terhadap ketepatan prediksi *rating* yang dihasilkan.

Adapun batasan masalah tugas akhir ini adalah sebagai berikut :

1. Metode yang digunakan dalam penentuan rekomendasi adalah *item-based collaborative filtering* yang dikembangkan dengan *item-based demographic data*.
2. *Dataset* yang direkomendasikan adalah *dataset* film yang berasal dari IMDB.com.
3. *Item* yang dijadikan objek rekomendasi adalah film yang telah di-*rating* oleh *user* lain.

1.3 Tujuan

Tujuan yang ingin dicapai dalam pengerjaan Tugas Akhir ini adalah sebagai berikut:

1. Mengimplementasikan pengembangan *collaborative filtering* dengan menggunakan *demographic data* pada *recommender system*.
2. Menganalisis pengaruh parameter α , β dan γ pada perhitungan *enhanced correlation* dan jumlah *neighborhood* pada proses *find nearest neighbor* untuk memperbaiki kualitas rekomendasi yang dihasilkan berdasarkan metode evaluasi MAE (*Mean Absolute Error*).
3. Membandingkan hasil prediksi antara metode *collaborative filtering* biasa dengan *demographic collaborative filtering* dengan cara menganalisis pengaruh *demographic data* berdasarkan hasil TOP-N *Recommendation*.

1.4 Metodologi penyelesaian masalah

Metodologi yang digunakan dalam memecahkan masalah di atas adalah dengan menggunakan langkah-langkah berikut:

1. Studi Literatur
Pencarian referensi dan sumber-sumber yang berhubungan dengan *recommender system*, *item-based collaborative filtering* dan *demographic data* dalam menyelesaikan tugas akhir ini.
2. Pengumpulan data
Mengumpulkan data-data produk yang diperlukan sebagai *training set* dan *test set*.
3. Perancangan Sistem
Perancangan terhadap *recommender system* yang akan dibangun.
4. Implementasi dan Pembangunan Sistem
Tahap pembangunan *recommender system* dengan melakukan pembangunan terhadap *database*, implementasi algoritma, kemudian

membuat *interface* aplikasi untuk memudahkan *user* dalam mengakses sistem.

5. Testing & Analisis

Melakukan pengujian dan analisis terhadap sistem yang dibangun. Melakukan analisis dari implementasi sistem dan pengujian hasil. *Mean Absolute Error* (MAE) digunakan untuk menghitung akurasi dari hasil rekomendasi. yang dihasilkan.

6. Pengambilan Kesimpulan dan penyusunan laporan Tugas Akhir.

Menyusun laporan tertulis berdasarkan hasil penelitian yang dilakukan dan memberikan kesimpulan mengenai hasil dari penelitian yang dilakukan, pemberian saran untuk pengembangan sistem yang dibangun ke depannya.



BAB V

Penutup

Pada bab ini disimpulkan hasil pengerjaan Tugas Akhir ini dan beberapa saran yang dapat mengembangkan Tugas Akhir ini.

5.1 Kesimpulan

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

1. Pada *dataset* IMDB, penjumlahan *similarity* antara *adjusted correlation* dan *demographic correlation* menjadikan prediksi menjadi semakin akurat. Hal ini disebabkan karena adanya penambahan *similarity* demografi yang menjadikan perhitungan prediksi menjadi lebih baik.
2. Nilai *neighborhood* optimal didapat ketika sebagian besar user merating item sebanyak nilai tersebut.
3. Hasil prediksi *rating* dengan menggunakan *collaborative filtering* yang digabungkan dengan *demographic data* lebih baik dibandingkan dengan tanpa penggabungan karena *demographic data* dari *item* lebih mencerminkan latar belakang dari *item* tersebut.
4. Besar *similarity* pada *item* yang telah di-*rating* sebelumnya mempengaruhi kecocokan *genre* antara *item* hasil rekomendasi dgn *item* sebelumnya. Dengan kata lain, semakin besar nilai *similarity neighborhood* maka akan semakin cocok pula *genre*-nya dengan *genre item* hasil rekomendasi tersebut.
5. Dengan menggunakan *demographic collaborative filtering*, prediksi *rating* yang dihasilkan tidak hanya mempertimbangkan *rating*-nya saja, tetapi juga melihat *genre* dari *item* yang telah di-*rating* sebelumnya. Hal ini dapat dibuktikan dengan adanya kecocokan *genre* antara *item* hasil rekomendasi dengan *item* sebelumnya.
6. Metoda *demographicCF* tidak menangani *sparsity problem* karena berdasarkan nilai MAE yang selalu bertambah besar disetiap *missing rate*-nya.

5.2 Saran

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

1. Untuk dapat memprediksi *rating* dalam jumlah dataset yang besar dan dalam waktu proses yang cepat, sebaiknya ditambahkan algoritma yang lebih efisien.
2. Agar hasil prediksi sesuai dengan latar belakang *user*, maka dapat menggunakan *user-based demographic data*.

Referensi

- [1] Ampazis, Nicholas. Collaborative Filtering Via Concept Decomposition On The Netflix Dataset. Diunduh pada : http://www.feeds2.com/netflix/Ampazis_ECAI08.pdf, 28 April 2010.
- [2] Amstrong, Scott, J., & Collopy, Fred. Error Measures For Generalizing About Forecasting Methods: Empirical Comparisons. Diunduh pada : <http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.36.9077&rep=rep1&type=pdf>, 5 Mei 2010.
- [3] Anne Yun-An Chen & Mcleod, Dennis. Collaborative Filtering for Information Recommendation Systems. Diunduh pada : <http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.100.7018&rep=rep1&type=pdf>, 22 Februari 2010.
- [4] Billsus, Daniel & Michael J. Pazzani. Learning Collaborative Information Filters. Diunduh pada : <http://www.aaai.org/Papers/Workshops/1998/WS-98-08/WS98-08-005.pdf>, 20 Februari 2010.
- [5] Burke, Robin. Hybrid Recommender Systems : Survey and Experiments. Diunduh pada : <http://gamejam.cti.depaul.edu/~rburke/pubs/burke-umuai02.pdf>, 17 Maret 2010.
- [6] E. Aimeur, G. Brassard, J. M. Fernandez, & F. S. Mani Onana. Privacy-Preserving Demographic Filtering. Diunduh pada : <http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.129.7348&rep=rep1&type=pdf>, 17 Maret 2010.
- [7] Kangas, Sonja. Collaborative Filtering and Recommendation Systems. Diunduh pada : <http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.120.1945&rep=rep1&type=pdf>, 2 Februari 2010.
- [8] Lyon, Cliffrod. *Movie* Recommender. Diunduh pada : <http://materialobjects.com/cf/MovieRecommender.pdf>, 7 Februari 2010.
- [9] Purwanto, Ari. Metode Analisis Rekomendasi Pada Sistem Rekomendasi (Contoh Kasus Pemanfaatan Pada Biro Wisata). Diunduh pada :

<http://aripurwanto.blog.upi.edu/files/2009/07/ilmiah.pdf>, 25 Maret 2010.

- [10] Sarwar, Badrul, Karypis, George, Konstan, Joseph & Reidl, Jhon. *Item-based Collaborative Filtering Recommendation Algorithms*. Diunduh pada : <http://portal.acm.org/citation.cfm?id=371920.372071>, 2 Februari 2010.
- [11] Vozalis, Emmanouil & Konstantinos G. Margaritis. *Analysis of Recommender System's Algorithm*. Diunduh pada : <http://users.uom.gr/~mans/papiria/hercma2003.pdf>, 15 Februari 2010.
- [12] Vozalis, Manolis & Konstantinos G. Margaritis. *Enhancing Collaborative Filtering With Demographic Data : The Case Of Item-based Filtering*. Diunduh pada : <http://users.uom.gr/~mans/papiria/voz-demog-ws-hu.pdf>, 16 Maret 2010.
- [13] Vozalis, Manolis & Konstantinos G. Margaritis. *Collaborative Filtering Enhanced By Demographic Correlation*. Diunduh pada : <http://eos.uom.gr/~mans/papiria/voz-demog-aiai.pdf>, 15 Maret 2010.