

## ANALISIS DAN IMPLEMENTASI ITEM-BASED CLUSTERING HYBRID METHOD RECOMMENDER SYSTEM

Ramadhanuz Awaludin Djamal<sup>1</sup>, Warih Maharani<sup>2</sup>, Angelina Prima Kurniati<sup>3</sup>

<sup>1</sup>Teknik Informatika, Fakultas Teknik Informatika, Universitas Telkom

---

### Abstrak

Recommender system adalah sebuah aplikasi yang mencari dan memberikan rekomendasi berupa item-item dengan memprediksi rating-ratingnya berdasarkan persamaan karakteristik user dalam memberikan informasi.

Tugas Akhir ini mengimplementasikan dan menganalisis Item-Based Clustering Hybrid Method (ICHM) yang merupakan recommender system yang menggabungkan pendekatan antara collaborative filtering dengan content based filtering. Penggabungan pendekatan content based filtering dengan collaborative filtering pada metode ICHM bertujuan untuk menanggulangi kekurangan yang ada pada kedua pendekatan sebelumnya.

Sistem rekomendasi dibangun berbasis web dengan menggunakan Java Servlet Pages (JSP) dan Apache Tomcat sebagai web server. Tugas akhir ini menganalisis akurasi prediksi rating yang dihasilkan oleh sistem rekomendasi dengan metode ICHM. Parameter yang digunakan dalam analisis adalah jumlah cluster dan koefisien  $c$  sebagai penggabungan nilai similarity.

Sistem rekomendasi dengan metode ICHM memiliki keunggulan dapat memprediksi item yang belum dirating sama sekali. Selain itu jumlah cluster mempengaruhi nilai similarity berdasarkan content item serta komposisi penggabungan berdasarkan koefisien  $c$  mempengaruhi hasil prediksi rating untuk active user. Akurasi prediksi yang dihasilkan oleh metode ICHM cenderung semakin meningkat dengan bertambahnya jumlah cluster.

**Kata Kunci :** recommender system, hybrid, metode ICHM, item-based clustering hybrid method

---

### Abstract

Recommender is an application that can search and recommend items by predicting their ratings based on the similarity of user's characteristic in giving the information.

In this final project, author implements Item-Based Clustering Hybrid Method (ICHM) which is one of hybrid recommender system that combines the collaborative filtering and content based filtering. The purpose of combination between content based filtering and collaborative filtering in ICHM for overcoming each filtering shortcomings.

The analysis carried out to the accuracy of rating prediction result given by the recommender system. The number of clusters and the value of coefficient  $c$  as variable in combining similarity value are used to be the parameters of analysis.

ICHM recommender system has an advantage that can predict new items that have no rating at all. Different number of clusters gives different value of similarity based on the item's content and the combination computation based on coefficient  $c$  affect the result of rating prediction. The increasing number of clusters in ICHM gives the prediction accuracy tend to increase.

**Keywords :** recommender system, hybrid, ICHM, item-based clustering hybrid method

---

# 1 Pendahuluan

## 1.1 Latar belakang masalah

*Recommender system* atau sistem rekomendasi dikembangkan untuk membantu user dalam memperoleh informasi yang dibutuhkan dalam internet akibat membanjirnya data yang tersedia, sehingga dapat memberikan keuntungan bagi pihak user maupun pengembang aplikasi Internet[1,12]. *Content-based filtering* dan *Collaborative filtering* adalah pendekatan yang paling umum untuk membangun sebuah sistem rekomendasi[1,7].

Pendekatan sistem rekomendasi dengan *content-based* bekerja dengan mencari item berdasarkan konten dari item-item yang telah dipilih atau diambil oleh user tersebut sebelumnya[1,4,7]. Beberapa penelitian telah menyebutkan bahwa pendekatan ini efektif, namun memiliki keterbatasan ketika seorang user meminta rekomendasi suatu item yang memiliki jenis konten yang berbeda dengan item-item yang pernah dipilih olehnya.

Untuk menutupi kekurangan *content-based filtering* tersebut, maka dibangun pendekatan *collaborative filtering* yang bekerja dengan mencari user yang memiliki selera yang sama terhadap item-item yang ada, misalkan cara *me-rating* film pada sistem rekomendasi film[1,7,8]. Pendekatan ini terbukti berhasil dalam beberapa penelitian dan praktis, namun pendekatan ini memiliki kelemahan di saat suatu item masuk dan sama sekali belum ada yang memberi rating, akibatnya item tersebut tidak akan pernah direkomendasikan ke user mana pun.

Oleh karena itu diperlukanlah pendekatan *hybrid* untuk menutupi kekurangan-kekurangan tersebut. Salah satunya adalah dengan memasukkan karakteristik *content-based filtering* ke dalam pendekatan *collaborative filtering* sehingga masalah masuknya item baru ke dalam database dapat ditangani.

*Item-based Clustering Hybrid Method* (ICHM) adalah salah satu metode yang menggunakan pendekatan *hybrid* atau menggabungkan kedua pendekatan tersebut. Kelebihan ICHM ini adalah dapat mengatasi masalah rekomendasi untuk *item* yang baru dan belum dirating[7]. ICHM membangun *Group-rating* berdasarkan konten atau atribut yang dimiliki *item* dan membagi item-item tersebut menjadi beberapa *cluster* atau grup. *Group-rating* ini yang meningkatkan performa dari *collaborative filtering* dalam fase perhitungan kemiripan.

Pada akhir implementasi, digunakan Mean Absolute Error (MAE) untuk mengevaluasi akurasi prediksi rating yang dihasilkan sistem rekomendasi dengan membandingkannya terhadap rating aktual yang diberikan oleh user[8,12]. MAE dipilih sebagai metrik evaluasi karena mudah dan banyak digunakan di beberapa penelitian.

## 1.2 Perumusan masalah

Permasalahan yang dibahas pada tugas akhir ini adalah sebagai berikut:

- Bagaimana menerapkan *recommender system* menggunakan ICHM.
- Bagaimana mengevaluasi hasil prediksi rating sistem rekomendasi dengan metode ICHM.

- c. Bagaimana penanganan prediksi untuk *item* baru yang belum *dirating* sama sekali menggunakan metode ICHM.

Batasan masalah pada tugas akhir ini adalah sebagai berikut:

- a. *Item* yang dijadikan objek rekomendasi adalah film yang telah *dirating* oleh *user* yang didapat dari *dataset* MovieLens.org.
- b. *Content item* yang digunakan adalah *genre* film.

### 1.3 Tujuan

Tujuan yang ingin dicapai dalam pembuatan Tugas Akhir ini antara lain :

- a. Membangun sebuah aplikasi yang menerapkan sistem rekomendasi dengan metode ICHM.
- b. Menganalisis dan mengukur tingkat *error* hasil prediksi rating dari sistem rekomendasi dengan menggunakan *Mean Absolute Error* (MAE) terhadap jumlah *cluster* dan koefisien *c* sebagai variabel penggabungan *similarity*.
- c. Menganalisis sistem rekomendasi dalam memprediksi *item* baru yang belum *dirating* sama sekali oleh *user*.

### 1.4 Metodologi penyelesaian masalah

- a. Studi Literatur :  
Pencarian referensi dan sumber pendukung lainnya yang dapat digunakan sebagai acuan dalam pembuatan tugas akhir ini.
- b. Analisis dan Desain :  
Menganalisis kebutuhan sistem dan merancang desain untuk *recommender system* yang akan dibangun.
- c. Implementasi sistem:  
Membangun aplikasi yang telah didefinisikan menjadi perangkat *recommender system*. Melakukan *preprocessing* pada matriks *item-content* yang digunakan untuk membangun matriks *group-rating* lalu menyimpannya ke dalam *database*.
- d. Analisis hasil implementasi:  
Skenario uji yang dilakukan adalah dengan menganalisis tingkat *error* dari hasil prediksi yang dihasilkan sistem yang sudah dibangun dengan menggunakan MAE terhadap rating aktualnya. Untuk skenario uji *item* baru dihapus semua *rating item* yang akan diuji dan kemudian dihitung tingkat *error* pada prediksi *rating* dengan menggunakan MAE.
- e. Penyusunan laporan  
Penyusunan laporan hasil penelitian yang telah dilakukan serta membuat kesimpulan dari hasil penelitian tersebut.

## 5 KESIMPULAN DAN SARAN

### 5.1 Kesimpulan

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

1. *Recommender system* dengan metode ICHM dapat memprediksi *item* baru yang belum pernah dirating sama sekali karena memperhitungkan *similarity* berdasarkan *genre item*.
2. Hasil prediksi tiap *item* berbeda bergantung kepada nilai *similarity* dan pola *user peratingan* user pada item-item sebelumnya.
3. Secara umum penambahan jumlah *cluster* hingga 70 buah cenderung meningkatkan akurasi prediksi baik untuk kasus *cold start* dan *non cold start problem*, namun akurasi turun pada jumlah *cluster* sebanyak 60 karena terdapat nilai *membership* yang saling bertolak belakang untuk beberapa *item* di beberapa *cluster*.
4. Secara umum penggabungan nilai *similarity* menghasilkan prediksi *rating* yang lebih akurat karena baik dengan perhitungan *collaborative filtering* atau *content based filtering* memiliki kekurangan masing-masing, dengan menggabungkan kedua pendekatan tersebut maka kekurangan yang ada dapat diatasi.

### 5.2 Saran

1. Menggunakan algoritma *clustering* dinamis agar tidak memerlukan input jumlah *cluster* yang diperlukan untuk membangun *group rating*.

## Daftar pustaka

- [1] Adomavicius, Gediminas and Tuzhilin, Alexander. 2005. *Toward the Next Generation of Recommender Systems: A Survey of the State-of-the-Art and Possible Extensions*. IEEE Transactions on Knowledge and Data Engineering, vol. 17, no. 6, June 2005.
- [2] Basu, Chumki., Hirsh, Hayim., and Cohen, William. 1998. *Recommendation as classification: Using social and content-based information in recommendation*. Proceedings of the Fifteenth National Conference on Artificial Intelligence, 714720.
- [3] Breese, Jhon S., Heckerman, David, and Kadie, Carl. 1998. *Empirical Analysis of Predictive Algorithms for Collaborative Filtering*. Microsoft Research, Microsoft Corporation.
- [4] Hauger, Stefan., Tso, Karen H. L., and Schmidt-Thieme, Lars. 2007. *Comparison of Recommender System Algorithms Focussing on the New-Item and User-Bias Problem*. Departement of Computer Science, University of Freiburg. Information System and Machine Learning Lab, University of Hildesheim.
- [5] Herlocker, Jonathan L. 2001. *Evaluating Collaborative Filtering Recommender System*. 2001 ACM 1073-0516/01/0300-0034.
- [6] Leben, Micheal. 2008. *Applying Item-based and User-based Collaborative Filtering on the Netflix Data*. Hasso-Plattner-Institut Potsdam.
- [7] Li, Qing and Kim, Byeong Man. 2002. *An Approach for Combining Content-based and Collaborative Filters*. Departement of Computer Sciences, Kumoh National Institute of Technology.
- [8] Nichols, David M., 1997. *Implicit Rating and Filtering*. Computing Department, Lancaster University, Lancaster, LA1 4YR. United Kingdom.
- [9] Sarwar, Badrul et al. 2001. *Item-based Collaborative Filtering Recommender System Algorithm*. GroupLens Research Group / Army HPC Research Center, Department of Computer Science and Engineering, University of Minnesota. Minneapolis.
- [10] Sarwar, Badrul M., et al. 2003. *Recommender Systems for Large-scale E-Commerce: Scalable Neighborhood Formation Using Clustering*. Department of Computer Science and Engineering University of Minnesota, Minneapolis, United States of America.
- [11] Vozalis, E. and Margaritis K. G. 1998. *Recommender Systems: An Experimental Comparison of two Filtering Algorithms*. University of Macedonia, GREECE.
- [12] Vozalis, E. and Margaritis, K. G. 2003. *Analysis of Recommender Systems Algorithms*.