

PERBANDINGAN METODE COLLABORATIVE FILTERING DAN HYBRID SEMANTIC SIMILARITY

Imam Fahrurrozi¹, Estu Muh Dwi Admoko², Anang Susilo³

Program Studi Komputer dan Sistem Informasi, Departemen Teknik Elektro dan Informatika,
Sekolah Vokasi, Universitas Gadjah Mada, Indonesia

Email: ¹imam.fahrurrozi@ugm.ac.id, ²estumuhdwi@ugm.ac.id, ³anangsusilo@ugm.ac.id

ABSTRACT

Recommender system is a component which has been developed for online commerce purposes. In this issue, one of the popular methods that has been widely used is collaborative filtering. However, this method has some drawbacks and needs to be improved. Therefore, in this research a combination of Collaborative Filtering (CF) and semantic similarity method has been compare with original CF, and the result expected reducing some deficiencies on the original collaborative filtering method. Based on the performance tests, the results conclude that the combination can reduce some weaknesses on the original collaborative filtering, especially on the cold-start item and sparsity issue.

Keywords: *recommender system, collaborative filtering, semantic similarity, combination, cold-start item, sparsity data.*

PENDAHULUAN

Transaksi perdagangan merupakan salah satu kegiatan yang sering kita jumpai pada kehidupan sehari-hari. Transaksi perdagangan tidak hanya menggunakan cara yang konvensional akan tetapi telah memanfaatkan teknologi informasi dan jaringan internet. Beberapa contoh transaksi online diantaranya adalah buku online, toko elektronik dan toko handphone, sudah banyak dimanfaatkan oleh pengguna dari berbagai tempat dan waktu.

Dengan banyaknya data yang ada di toko online, maka pengguna akan kesulitan untuk memilih produk yang diinginkan oleh pengguna tersebut, dengan sedikitnya waktu pengguna untuk memilih produk yang tepat atau yang pengguna inginkan. Beberapa peneliti telah banyak mengusulkan beberapa cara untuk meningkatkan kinerja system aplikasi online tersebut sedemikian rupa semakin mudah, cepat untuk digunakan bagi pengguna. Misalnya sistem-sistem online yang dikembangkan dengan menggunakan teknik-teknik komputasi cerdas (Wanarsup *et al.*, 2008; Sun dan Zhao, 2009), data mining (Liu *et al.*, 2009), prediksi dan rekomendasi (Sarwar *et al.*, 2001).

Metode *collaborative filtering* (CF) telah banyak dipakai pada sistem rekomendasi perdagangan online (Shambour dan Lu, 2011). Metode CF *pada prinsipnya adalah proses penyaringan data berdasarkan kemiripan informasi, karakteristik atau profil dari para pembeli.*

Menurut Montiel dan Montes (2009) metode *collaborative filtering* masih mempunyai kelemahan *sparsity* (keterbatasan) dan *the cold-start item*. *Sparsity* akan terjadi jika *rating* yang didapat dari *user* dengan penyebaran yang tidak merata dan pasangan *user* yang memiliki *rating item* yang sama sangat sedikit sehingga tidak memenuhi kuota untuk penghitungan nilai similaritas. Sedangkan *cold-start item* mempengaruhi *item* yang tidak mempunyai *rating* atau mempunyai sedikit *rating*, dengan demikian, sistem tidak bisa tepat menemukan tetangga yang *similar* untuk *item* baru (Sarwar *et al.*, 2001; Adomavicius dan Tuzhilin, 2005; Schafer *et al.*, 2007).

Kesamaan dari produk-produk yang berbeda, dapat dispesifikasikan dalam bentuk informasi semantik yaitu kesamaan semantic dari beberapa item yang didasarkan pada sumber pengetahuan (*ontology*). Dalam konteks sistem atau aplikasi rekomendasi online di atas dapat juga dikombinasikan dengan menambahkan *semantic information* dari *item*, yang berguna untuk memberikan informasi sebagai model rekomendasi berdasarkan sumber pengetahuan (Shambour dan Lu, 2011). Berdasarkan permasalahan pada metode CF, penelitian ini mencoba membandingkan antara *hybrid semantic similarity* (CF dan *leacock chodorow*) dan *traditional collaborative filtering* sehingga didapatkan mana metode yang paling baik tingkat prediksi rekomendasinya dengan menggunakan pengujian *Mean Absolute Error* (MAE).

BAHAN DAN METODE PENELITIAN

2.1 Data Penelitian

Data diperlukan untuk menghasilkan informasi yang baik, karena informasi pada dasarnya merupakan hasil dari pengolahan data yang diinputkan pada sistem. Pada sistem rekomendasi ini, sumber data berasal dari data internal, yang berasal dari dalam organisasi untuk mendukung sistem rekomendasi yang akan dirancang. Adapun beberapa data internal yang dibutuhkan adalah data item aksesoris *handphone*, data spesifikasi dan data kategori aksesoris *handphone*. Yang kedua berasal dari data eksternal yang berasal dari luar organisasi namun tetap memiliki pengaruh dalam menciptakan sistem rekomendasi yang baik. Data eksternal yang dibutuhkan adalah data pengguna sistem, data harga, data jarak semantik antar aksesoris *handphone* yang berasal dari *wordnet* dan data rating pengguna ke aksesoris *handphone*. Data eksternal diambil langsung pada toko aksesoris *handphone*.

2.2 Collaborative Filtering (CF)

Collaborative filtering merupakan proses penyaringan atau pengevaluasian item menggunakan opini dari orang lain (Schafer *et al.*, 2007). *Collaborative filtering*

menghasilkan prediksi atau rekomendasi bagi pengguna / pelanggan yang dituju terhadap satu *item* atau lebih. Sarwar et al menjelaskan rumus (1) untuk menghitung similaritas dengan metode *collaborative filtering* (Sarwar et al., 2001), sebagai berikut :

$$Sim(i, j) = \frac{\sum_{u \in U} (R_{u,i} - \bar{R}_i)(R_{u,i} - \bar{R}_j)}{\sqrt{\sum_{u \in U} (R_{u,i} - \bar{R}_i)^2} \sqrt{\sum_{u \in U} (R_{u,i} - \bar{R}_j)^2}} \quad (1)$$

Keterangan :

Sim(i,j)= Nilai kemiripan antara aksesoris i dan aksesoris j

$\sum_{u \in U}$ = Penjumlahan dari *user* pertama hingga *user* terakhir dari semua anggota *user*

\bar{R}_i = Nilai rata-rata rating yang diberikan pengguna u ke aksesoris i

\bar{R}_j = Nilai rata-rata rating yang diberikan pengguna u ke aksesoris j

$R_{u,i}$ = Nilai rating yang diberikan pengguna u pada aksesoris i

$R_{u,j}$ = Nilai rating yang diberikan pengguna u pada aksesoris j

Hasil dari metode CF ini akan dihasilkan data-data kesamaan antar produk yang ada Di Sistem yang kemudian akan dibandingkan dengan metode Hybrid antara CF dan Semantic Similarity.

2.3 Similaritas Semantik dengan Metode Leacock Chodorow

Similaritas semantik memiliki peran penting dalam *Information Retrieval*, integrasi informasi, sistem rekomendasi dan lainnya yang mencakup perbandingan antar konsep. Similaritas semantik dapat didefinisikan sebagai tingkat pengukuran set istilah atau kata pada sumber informasi yang berdasarkan tingkat kesamaan konten semantik pada kumpulan set istilah tersebut (Saruladha, 2011).

Pederson et al. (2007) mendefinisikan similaritas semantik sebagai sebuah tolak ukur tingkat kesamaan atau keterkaitan antara dua buah konsep. Sedangkan Russell dan Norvig (2003) mendefinisikan similaritas semantik sebagai pengukuran secara kuantitas fitur intrinsik yang dimiliki oleh dua konsep.

Penelitian ini menggunakan perhitungan similaritas semantik berbasis jarak karena mempunyai performa perhitungan yang baik (Ganesan et al.,2003). Sehingga metode yang akan digunakan adalah metode similaritas semantik Leacock Chodorow (LCH) yang ditunjukkan dengan rumus (2) sebagai berikut :

$$SemSim(i, j) = \log\left(\frac{2 \cdot D}{length(i, j)}\right) \quad (2)$$

Keterangan :

i = aksesoris i

j = aksesoris j

length (i,j) = panjang lintasan antara dua aksesoris

D = maksimum *depth* (kedalaman) dari *node* ke *root* antara dua aksesoris

2.3.1 Metode Hybrid Semantic Similarity (Collaborative Filtering dan Leacock Chodorow)

Tahap ini mengintegrasikan dua similaritas yaitu similaritas semantik dan collaborative filtering untuk mendapatkan hybrid semantic similarity (Guo dan Lu, 2005). Sehingga yang digunakan adalah rumus (3) sebagai berikut :

$$\text{CombinedSim}(i,j) = \alpha * \text{Sim}(i,j) + (1 - \alpha) * \text{SemSim}(i,j) \quad (3)$$

Keterangan:

- $\text{Sim}(i,j)$ = similaritas dengan collaborative filtering
- $\text{SemSim}(i,j)$ = similaritas semantik dengan leacock chodorow
- α = parameter kombinasi semantik yang menunjukkan bobot dari similaritas semantik.

2.3.2 Perhitungan Prediksi dan Rekomendasi

Pada perhitungan prediksi rating pada suatu produk dapat dibagi menjadi dua perhitungan berdasarkan kondisi, yaitu ketika kondisi aksesoris *handphone* yang belum pernah dirating atau nilai rata-rata rating yang diberikan pengguna ke aksesoris *handphone* bernilai nol maka menggunakan metode *weighted sum* untuk memprediksi nilai *rating* (Djamal *et al.*, 2010; Zhang *et al.*, 2011) dengan rumus (4) sebagai berikut :

$$\text{prediction}_{u,j} = \frac{\sum_{u \in U} \text{CombinedSim}(i,j) \times R_{ui}}{(\sum_{u \in U} |\text{CombinedSim}(i,j)|)} \quad (4)$$

Keterangan:

- $\sum_{u \in U}$ = Penjumlahan dari *user* pertama hingga *user* terakhir dari semua anggota *user*
- i,j = aksesoris *handphone* i dan j
- R_{ui} = rating yang diberikan pengguna u ke aksesoris *handphone* i

Sedangkan untuk aksesoris *handphone* yang sudah pernah dirating oleh pengguna atau aksesoris *handphone* yang telah mempunyai nilai rata-rata rating yang diberikan pengguna ke aksesoris *handphone*, maka digunakan metode *weighted average of deviation* untuk memprediksi nilai *rating* (Djamal *et al.*, 2010) dengan rumus (5) sebagai berikut :

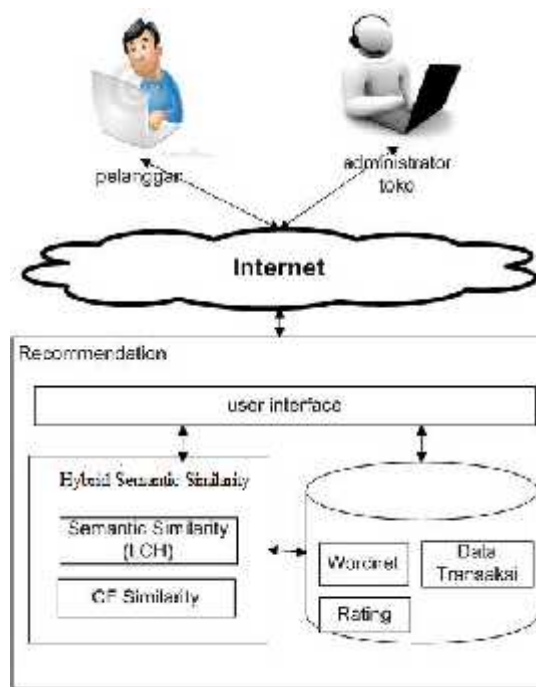
$$\text{prediction}_{u,j} = \bar{R}_j + \frac{\sum_{u \in U} \text{CombinedSim}(i,j) \times (R_{ui} - \bar{R}_i)}{(\sum_{u \in U} |\text{CombinedSim}(i,j)|)} \quad (5)$$

Keterangan:

- $\sum_{u \in U}$ = Penjumlahan dari *user* pertama hingga *user* terakhir dari semua anggota *user*
- i,j = aksesoris *handphone* i dan j
- R_{ui} = rating yang diberikan pengguna ke aksesoris
- \bar{R}_i, \bar{R}_j = rata-rata rating pengguna ke aksesoris *handphone* i dan j

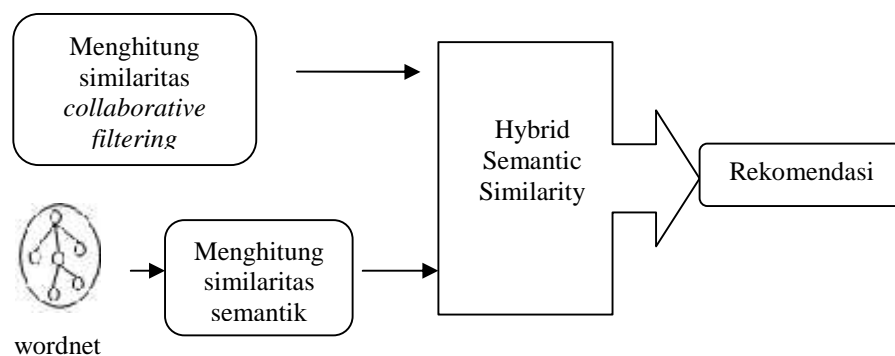
2.4 Rancangan Model Sistem

Secara umum arsitektur sistem yang diajukan pada penelitian ini disajikan pada Gambar 1. Sistem rekomendasi yang akan dibangun mempunyai dua pengguna yaitu pelanggan dan administrator toko. Pelanggan melakukan aktifitas pembelian dan pemberian nilai (*rating*) terhadap aksesoris handphne, sedangkan administrator toko mengupdate data yang ada di toko online. *User interface* sistem mewajibkan para pengguna untuk *login* terlebih dahulu sebelum masuk ke dalam sistem.



Gambar 1. Arsitektur Sistem

Gambar 1 memperlihatkan proses rekomendasi terdiri dari metode *Leacock Chodorow* (LCH) dan *Collaborative Filtering* (CF). Perhitungan data dari kedua metode berasal dari data wordnet, rating dan transaksi yang berasal dari *database*. Tahapan untuk menghasilkan rekomendasi berbasis semantik dan *collaborative filtering* secara lebih detail juga diperlihatkan pada Gambar 2.



Gambar 2. Tahapan Untuk Menghasilkan Rekomendasi

Hasil dari Tahapan pada Gambar 2 kemudian dibandingkan dengan hasil dari Rekomendasi yang dihasilkan metode traditional collaborative filtering yang dapat dilihat pada bagian Hasil dan Pembahasan.

2.5 Rancangan Pengujian

Penelitian ini menggunakan *metric standart Mean Absolute Error (MAE)* pada pengujiannya. Metrik mean absolute error atau MAE digunakan untuk menghitung tingkat akurasi atau besar error hasil prediksi rating dari sistem terhadap rating sebenarnya yang pengguna berikan terhadap suatu *item* (Herlocker et. al, 2004; Vozalis dan Margaritis, 2003).

MAE diperoleh dengan menghitung error absolut dari N pasang rating asli dan prediksi, kemudian menghitung rata-ratanya, seperti pada rumus (6) sebagai berikut :

$$MAE = \frac{\sum_{i=1}^n |q_i - p_i|}{n} \quad (6)$$

Keterangan :- p_i = rating yang diprediksi
- q_i = rating yang sebenarnya

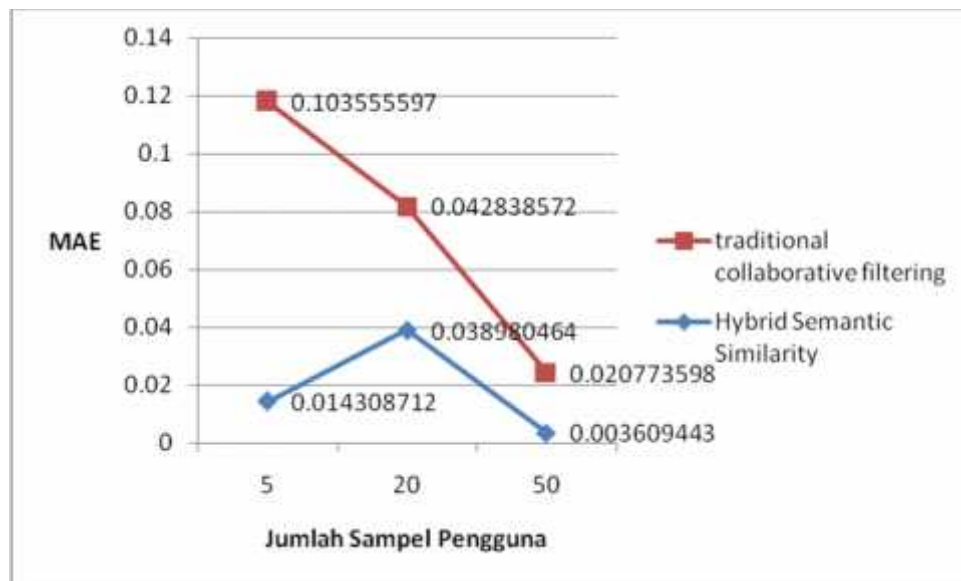
Rating yang diprediksi berasal dari output prediksi dari sistem dan rating yang sebenarnya berasal dari hasil kuisioner dari pengguna sistem. Data yang digunakan pada *prototype* toko aksesoris *handphone* dengan menggunakan metode yang diusulkan akan dibandingkan dengan *traditional collaborative filtering*. Perbandingan hasil rekomendasi akan dilakukan oleh 10 sampai 50 pengguna.

HASIL DAN PEMBAHASAN

Pada bagian ini kami membandingkan hasil penelitian dari metode *hybrid semantic similarity* dan *traditional collaborative filtering* berdasarkan nilai *Mean Absolute Error (MAE)*. Perbandingan ini memperlihatkan perbandingan nilai MAE antara metode *hybrid semantic similarity* dengan tradisional *collaborative filtering* yang dipengaruhi oleh jumlah sampel pengguna dan jumlah data koleksi aksesoris *handphone*. Semakin kecil nilai MAE maka semakin bagus nilai prediksi dan rekomendasi yang diberikan oleh sistem.

Perbandingan ini menggunakan keragaman jumlah data dan hasil melalui 3 skenario perbandingan yaitu 5 pengguna 5 data produk aksesoris *handphone*, 20 pengguna 15 data produk aksesoris *handphone* dan 50 pengguna 36 data produk aksesoris *handphone*. Berdasarkan Gambar 3, nilai MAE metode *hybrid semantic similarity* relatif lebih kecil jika dibandingkan dengan nilai MAE dari metode tradisional *collaborative filtering*, hal ini menunjukkan bahwa metode *hybrid semantic similarity* memiliki prediksi dan rekomendasi yang lebih bagus daripada metode tradisional *collaborative filtering*. Dari Gambar 3 menunjukkan nilai MAE metode *hybrid semantic similarity* mempunyai nilai terkecil yaitu

0,0036 ketika dibandingkan dengan tradisional *collaborative filtering standart* (CF) dengan nilai MAE = 0,0207.



Gambar 3. Perbandingan Akurasi Hasil Prediksi Rating Pada Metode *Hybrid Semantic Similarity* Dengan Tradisional *Collaborative Filtering* Berdasarkan Jumlah *Sample Pengguna*

KESIMPULAN

Berdasarkan penelitian dan pembahasan yang telah dilakukan, diperoleh kesimpulan bahwa metode *hybrid semantic similarity* (metode *leacock chodorow*) mempunyai tingkat kesalahan yang lebih sedikit dan mempunyai tingkat prediksi rekomendasi yang lebih baik daripada tradisional *collaborative filtering* yang dibuktikan dengan pengujian yang dilakukan di bab hasil dan pembahasan.

UCAPAN TERIMA KASIH

Kami Ucapkan Terimakasih kepada Tim PPM Sekolah Vokasi Universitas Gadjah Mada yang telah mensupport dan membiayai penelitian yang dilakukan penulis.

DAFTAR PUSTAKA

Adomavicius, G., & Tuzhilin, A. (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, pp. 734-749.

Djamal, R. A., Maharani, W., & Kurniati, A.P. (2010). Analisis dan Implementasi Metode Item-Based Clustering Hybrid pada Recommender System, *Konferensi Nasional Sistem dan Informatika*, pp. 216-222.

- Ganesan, P., Garcia-Molina, H., & Widom, J. (2003) Exploiting Hierarchical Domain Structure to Compute Similarity, *ACM Transactions on Information Systems (TOIS)*, Vol. 21, No.1, pp.64-93.
- Guo, X., & Lu, J. (2005). Recommending Trade Exhibitions by Integrating Semantic Information with Collaborative Filtering, *Web Intelligence Proceedings IEEE/WIC/ACM International Conference*, pp. 747-750.
- Herlocker, J. L., Konstan, J. A., Terveen, L. G., & Riedl, J. T. (2004). Evaluating Collaborative Filtering Recommender Systems, *ACM Transactions on Information Systems (TOIS)*, Vol 22, No.1, pp. 5-53.
- Liu, X., Jia, S., Liu, E., & Zhang, Z. (2009). Application of Web-based Data Mining in Personalized Online Recruiting System, *International Conference on Management and Service Science, 2009. MASS '09*, pp. 1-4.
- Montiel, R.M., & Montes, A.J. F. (2009). Semantically Enhanced Recommender Systems., *On the Move to Meaningful Internet Systems: OTM 2009 Workshops Springer Berlin Heidelberg*, pp. 604-609.
- Pedersen T., Pakhomov S. and Patwardhan S. (2007). Measures of Semantic Similarity and Relatedness in the Medical Domain, University of Minnesota Digital Technology Center Research Report DTC 2005/12, Vol. 40, No. 3.
- Russell S. and Norvig P. (2003) Artificial intelligence: A Modern Approach 2nd Edition, New Jersey, Prentice Hall.
- Saruladha, K. (2011). Semantic Similarity Measures for Information Retrieval Systems Using Ontology, Thesis, Department Computer Science Pondicherry University, India.
- Sarwar, B., Karypis, G., Konstan, J., & Riedl, J. (2001). Item-based Collaborative Filtering Recommendation Algorithms. In *Proceedings of the 10th international conference on World Wide Web*, pp. 285-295.
- Schafer, J. B., Frankowski, D., Herlocker, J., & Sen, S. (2007). Collaborative filtering Recommender Systems. , *In The adaptive web, Springer Berlin Heidelberg* . , pp. 291-324.
- Shambour, Q. & Lu, J. (2011). A Hybrid Multi-Criteria Semantic-enhanced Collaborative Filtering Approach for Personalized Recommendations, *IEEE/WIC/ACM International Conferences on Web Intelligence and Intelligent Agent Technology* , Vol. 1, pp. 71-78.
- Sun, X., & Zhao, W. (2009). Design and Implementation of an E-learning Model Based on WUM Techniques, *International Conference on E-Learning, E-Business, Enterprise Information Systems, and E-Government, 2009. EEEE '09.*, pp. 248n-251.
- Vozalis, E., & Margaritis, K. G. (2003). Analysis of Recommender Systems Algorithms, *Proceedings of the 6th Hellenic European Conference on Computer Mathematics and its Applications* , Athens, Greece.

Wanarsup, W., Pattamavorakun, Sn., & Pattamavorakun, St., Intelligent Personalization Job Web Site, *Ninth ACIS International Conference on Software Engineering, Artificial Intelligence, Networking, and Parallel/Distributed Computing, 2008. SNPD '08*, pp. 959 – 964.

Zhang, L., Zhang, X., Chen, Q., Zhu, Z., & Shi, Y. (2011). Domain-Knowledge Driven Recommendation Method and Its Application, *Computational Sciences and Optimization (CSO) IEEE Fourth International Joint Conference*, pp. 21-25.