

PENENTUAN REKOMENDASI PRODUK MENGGUNAKAN METODE COLLABORATIVE FILTERING

OLEH
MOH FARID EFENDI(1515015094)

JURUSAN TEKNIK INFORMATIKA,
FAKULTAS TEKNOLOGI INFORMASI DAN KOMUNIKASI,
UNIVERSITAS MULAWARMAN
ALAMAT : JL.BARONG TONGKOK, KAMPUS GUNUNG KELUA, SAMARINDA
75123

Email: efendyfarid@gmail.com

ABSTRAK

Penulisan Jurnal ini bertujuan untuk menebus kesalahan karena absen pada perkuliahan praktikum ADPL serta menjelaskan tentang Sistem Penentuan Rekomendasi Produk Menggunakan Metode Collaborative Filtering yang dapat digunakan untuk menentukan ketertarikan seseorang pada sesuatu. Metode yang digunakan dalam sistem ini adalah metode Collaborative Filtering yang banyak digunakan oleh situs periklanan(e-commerce) untuk memprediksi keinginan viewer dan menawarkan produk kepada viewer, hasil dari jurnal adalah analisis dari metode yang mungkin dapat di aplikasikan pada suatu instansi yang membutuhkan untuk menggunakan e-commerce.

Kata Kunci: Collaborative Filtering, system E-commerce, viewer interest, prediction method.

seperti buku, music dalam film. Sasarannya adalah pengunjung toko online, di mana system ini memberikan prediksi dan penawaran terbaik produk apa yang disukai. Data referensi produk dari pengunjung disimpan untuk kemudian diolah menggunakan teknik statistik sehingga menghasilkan rekomendasi barang terkait[1]. Teknologi ini banyak diadopsi oleh perusahaan besar di dunia e-commerce dalam seperti Amazon, Netflix dan Pandora. Sedemikian pentingnya recomender system ini bagi ecommerce karena kemampuannya untuk memberikan referensi yang disukai oleh customer, membuat strategi marketing, hingga meningkatkan loyalitas pelanggan. Bahkan kompetisi pun telah diadakan oleh Netflix untuk menguji hasil rekomendasi dengan nilai yang cukup besar. Riset di recommender system menjadi begitu penting karena perannya di bidang e-commerce telah terbukti banyak memberikan terobosan baru di dunia bisnis dan ilmu pengetahuan.

1. PENDAHULUAN

Recommendation system banyak digunakan untuk membuat prediksi

2. COLLABORATIVE FILTERING

merupakan salah satu algoritma yang digunakan untuk menyusun recommender system dan telah terbukti memberikan hasil yang sangat baik. Rating produk merupakan elemen terpenting dari algoritma ini, rating diperoleh dari sebagian besar customer di mana customer secara explicit memberikan penilaiannya terhadap produk. Kesimpulannya ialah system memberikan imbal balik kepada customer dengan mengolah data-data tersebut, sebagai gambaran dari skala nol sampai 5 yang mengindikasikan penilaian yang paling tidak disukai hingga paling disukai menurut sudut pandang customer, data ini memungkinkan untuk dilakukannya perhitungan statistik yang hasilnya menunjukkan produk mana yang diberikan rating tinggi oleh customer. Collaborative filtering menggunakan database yang diperoleh dari user. Ada dua komponen utama dalam data ini agar dapat membuat prediksi bagi recommender system yaitu user dan item. Keduanya membentuk rating matrix berupa m user $\{u_1, u_2, u_3, \dots, u_m\}$ dan daftar n item $\{i_1, i_2, i_3, \dots, i_n\}$. Di mana setiap user memberikan penilaiannya pada item berupa rating dalam skala 1 sampai 5. Rating ini dilambangkan dengan $I_{u,i}$. Tidak semua user memberikan rating ke setiap produk karena berbagai macam faktor, hal ini menyebabkan banyaknya missing value yang mengakibatkan sparsity pada data. User item rating matrix dapat digambarkan dengan table di bawah.

Tabel 1 Rating Matrix

	i_1	i_2	i_3	i_4	i_m
u_1	1	...	3	...	
u_2	5	4	
u_3	...	5	3	...	
u_4	...	4	
u_m					

Collaborative filtering dibagi menjadi dua kategori[3] yaitu memory based, model based dan gabungan keduanya menjadi hybrid recommendation system[4] bertujuan untuk mengatasi kelemahan yang muncul pada kedua kategori sebelumnya. Memory based recommendation menggunakan user rating sebagai bahan untuk menemukan similarity atau derajat kesamaan antar user. Di domain bisnis algoritma ini telah diterapkan pada situs Amazon, keunggulannya adalah kemudahan dalam implementasi dan sangat efektif. Sedangkan untuk model based recommendation tidak jauh berbeda dengan algoritma sebelumnya, yaitu menggunakan rating sebagai sumber data. Namun algoritma ini menggunakan teknik-teknik di data mining atau machine learning seperti Bayesian dan clustering. Gabungan dari model dan memory based membentuk hybrid recommender system. Algoritma ini diciptakan untuk mengatasi kelemahan yang terdapat pada dua algoritma sebelumnya seperti sparsity. Berikut adalah perbandingan antara memory based dan model based recommendation system[5]: A. Memory based Teknik yang dipakai adalah: Neighbor-based CF (item based/user-based CF. Algorithms with Pearson/vector cosine correlation) Item-based/user-based top-N

Recommendations. Keunggulannya berupa implementasi mudah, mudah menambahkan data-data baru tidak perlu mempertimbangkan content item yang direkomendasikan, skala yang baik dengan co-rated item. Sedangkan kekurangan dari algoritma ini adalah bergantung pada rating dari user, menurunnya performa jika data jarang, skalabilitas yang terbatas pada dataset yang besar. Model Based Teknik yang dipakai adalah: Bayesian belief nets CF, clustering CF, MDP-based CF, latent semantic CF, sparse factor analysis, CF using dimensionality, reduction techniques seperti SVD dan PC. Kelebihan: dapat mengatasi masalah data yang jarang, skalability dan masalah lainnya, akurasi meningkat, memberikan intuitive rational untuk rekomendasi. Kekurangannya adalah diperlukannya sumber daya yang besar untuk proses komputasi. C. Hybrid Recommendation System Teknik yang digunakan berupa content based CF (fab), content boosted CF, hybrid CF kombinasi memory based dan model based (personality diagnosis). Teknik ini sebagai solusi atas kelemahan yang terdapat pada memory dan model based CF seperti sparsity dan grayship, sehingga meningkatkan akurasi prediksi. Beberapa kelemahan juga muncul seperti meningkatnya kompleksitas program hingga perlunya data external yang terkadang tidak tersedia.

3. EVALUASI RECOMMENDER SYSTEM

Efektifitas dan performa suatu algoritma rekomendasi dapat dilihat dari hasil evaluasi. Tipe evaluasi yang dipakai sangat dipengaruhi oleh jenis recommender system yang dipakai. Evaluasi untuk recommender system

tidak jauh berbeda dengan teknik yang digunakan dalam machine learning dan information retrieval. Pengujian ini dapat diklasifikasikan menjadi tiga jenis[6] yaitu predictive accuracy metric seperti Mean Absolute Error(MAE) dan variasinya, classification accuracy metrics seperti precision, recall, F1-measure dan ROC sensitivity; dan tipe klasifikasi yang ketiga adalah rank accuracy metric seperti Pearson's product-moment correlation, Kendall's Tau, Mean Average Precision(MAP), half-life utility dan normalized distance performance Metric(NDPM). Tidak semua jenis evaluasi ini diterapkan oleh aplikasi karena karakteristik dan tujuan tiap aplikasi berbeda. Untuk jenis evaluasi apa saja yang diuji cobakan oleh tiap aplikasi akan dibahas di bab berikutnya. Evaluasi dilakukan dengan cara memotong dataset menjadi dua: train dan evaluasi. Skema pengujian ini umum dilakukan dalam pengujian machine learning. Pada umumnya peneliti menggunakan real dataset seperti movielens, jester dan Netflix. Aplikasi open source recommender system selain melakukan testing juga telah disertakan mekanisme evaluasinya. Hal ini memudahkan pengembangan algoritma baru di atas platform aplikasi tersebut untuk diuji coba seberapa besar performanya.

4. KENDALA DALAM RECOMMENDER SYSTEM

A. Sparsity Sparsity diidentikan dengan data yang jarang atau belum terisi penuh. Hal ini diakibatkan oleh adanya item baru yang muncul, Item baru yang muncul bisa jadi kurang sesuai dengan keinginan user karena kurangnya referensi data. User rating merupakan bahan mentah dalam menghasilkan rekomendasi. Item yang tidak mendapatkan rating akan sulit untuk

ditemukan kaitannya dengan item yang lain. Penyebab lain adalah user baru biasanya tidak memberikan ratingnya kepada suatu produk. Bisa jadi karena user tidak mengetahui barang yang dia suka atau membiarkan begitu saja produk yang ditampilkan tanpa memberikan rating. Masih banyak lagi factor yang menyebabkan data sparsity ini sehingga menjadi tantangan terbesar dalam collaborative filtering dan paling sulit untuk diatasi[7]. Beberapa riset telah dilakukan untuk mengatasi masalah ini. Berikut adalah tabel riset sebelumnya untuk mengatasi data sparsity.

Tabel 2 Riset pada data Sparsity

No.	Penelitian	Teknik
1	[8]	collaborative filtering using cluster-based
2	[9]	Item Classification
3	[10]	Case-based Reasoning and Item-based Collaborative Filtering
4	[11]	BP Neural network

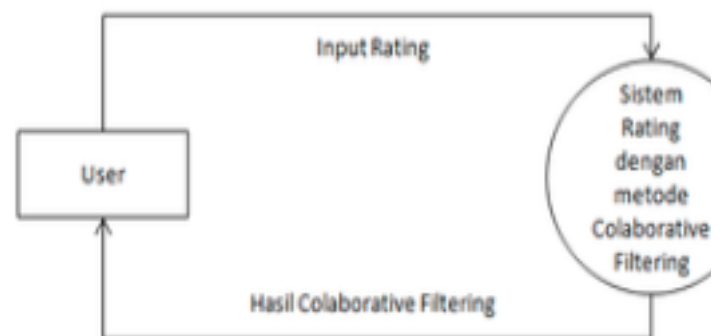
B. Scalability Collaborative filtering memerlukan training data yang semakin besar agar produk yang direkomendasikan akurat. Amazon, sebagai salah satu situs e-commerce terbesar, semakin hari data yang tersimpan akan semakin besar seiring pertumbuhan jumlah pengunjung. Hal ini tentu saja memakan banyak sumber daya dan komputasi. Komputasi untuk data dalam skala besar menjadi syarat penting untuk rekomen system dalam dunia bisnis dan riset. Oleh karena itu selain spesifikasi hardware, algoritma yang dipakai juga turut berperan performa rekomen system. Tabel berikut adalah riset dalam collaborative filtering untuk mengatasi masalah di skalabilitas ini

5. METODOLOGI

5.1.ANALISA KEBUTUHAN

Secara umum, sistem rekomendasi bertugas membantu para user dengan cara menyediakan rekomendasi bagi para usernya. Rekomendasi yang diberikan berupa *produk - produk*. Pada sistem ini user dapat memberikan *rating* pada berbagai macam produk yang menurutnya memiliki spesifikasi yang bagus dan layak menjadi *produk* terfavorit. *Rating* yang diberikan oleh masing-masing user digunakan sebagai dasar untuk menentukan rekomendasi *produk*.

5.2. DIAGRAM KONTEKS



Gambar 1. Diagram Konteks

Diagram konteks diatas menunjukkan bahwa user sebagai entitas luar(*Terminator*) menginputkan rating produk yang telah di lihat atau dipakai. Proses Sistem Rating dengan metode Collaborative filtering akan memproses hasil dari inputan user sehingga menghasilkan rekomendasi barang yang sesuai terhadap data yang telah di proses dan di filter.

5.3PERHITUNGAN NILAI KEMIRIPAN

Tahap ini adalah tahapan untuk

membuat nilai kemiripan diantara produk yang telah *dirating* oleh *user*. Untuk membuat nilai kemiripan, digunakan persamaan *adjusted-cosine* (2) yaitu :

$$sim(i,j) = \frac{\sum_{u \in U} (R_{u,i} - \bar{R}_u)(R_{u,j} - \bar{R}_u)}{\sqrt{\sum_{u \in U} (R_{u,i} - \bar{R}_u)^2} \sqrt{\sum_{u \in U} (R_{u,j} - \bar{R}_u)^2}} \dots (2)$$

$sim(i,j)$ = Nilai kemiripan antara item i dan item j
 $u \in U$ = Himpunan user u yang merating item i dan item
 $R_{u,i}$ = Rating user u pada item i
 $R_{u,j}$ = Rating user u pada item j
 \bar{R}_u = Nilai rata-rata rating user u

Dalam sistem yang dikembangkan, nilai kemiripan ini tidak disimpan dalam database, tetapi dikalkulasi ketika akan melakukan rekomendasi dan hasilnya disimpan dalam suatu array. Hal tersebut dilakukan karena nilai kemiripan cenderung terus berubah ketika ada rating baru yang masuk.

Efek yang diakibatkan oleh hal tersebut adalah cost komputasi dan waktu akan meningkat dalam setiap melakukan proses prediksi, tetapi dilain sisi prediksi sistem akan lebih akurat karena data kemiripan yang digunakan dalam dalam proses prediksi dihasilkan secara langsung saat itu juga.

5.4. PENGHITUNGAN PREDIKSI

Sesudah nilai kemiripan antar game didapatkan, maka tahap selanjutnya adalah membuat prediksi *rating* terhadap game yang belum *dirating* oleh *user*. Untuk menghitung prediksi *rating*, digunakan persamaan 3 yaitu persamaan *weighted sum* dengan menerapkan *nearest neighbor* dengan menentukan jumlah *neighbor* yang digunakan dalam proses penghitung prediksi. berikut adalah persamaan *weighed sum* :

$$P(u,j) = \frac{\sum_{i \in I_u^K(j)} (R_{u,i} * S_{i,j})}{\sum_{i \in I_u^K(j)} |S_{i,j}|} \dots \dots \dots$$

Keterangan :
 $P(u,j)$ = Prediksi untuk user u pada item j
 $i \in I_u^K(j)$ = Himpunan K item yang mirip dengan item j
 $R_{u,i}$ = Rating user u pada item i
 $S_{i,j}$ = Nilai kemiripan antara item i dan item j

Dalam proses memprediksi suatu *rating* produk terhadap seorang *user*, sebelumnya ditentukan dahulu berapa jumlah K yang akan digunakan. K dalam hal ini merupakan sejumlah produk yang telah *dirating* oleh *user* dan mempunyai nilai kemiripan paling tinggi dengan produk yang akan diprediksi *rating*nya.

5.5. MENGHITUNG NILAI MAE(MEAN ABSOLUTE ERROR)

Proses ini dilakukan untuk menghitung keakuratan sistem dalam melakukan prediksi *rating*. Persamaan yang digunakan adalah *mean absolute error*

(persamaan 4) yaitu:

$$MAE = \frac{\sum_{i=1}^N |p_i - q_i|}{N} \dots\dots\dots$$

Keterangan :

MAE = Nilai rata-rata kesalahan hitung

N = Jumlah item yang dihitung

p_i = Nilai prediksi item ke- i

q_i = Nilai rating sebenarnya item ke- i

Menghitung nilai absolut selisih *rating* yang ada pada tabel hasil prediksi ` dengan *rating* sebenarnya dibagi dengan jumlah data *rating*.

Kesimpulan

Dari penggunaan dan metode diatas dapat diambil kesimpulan bahwa sistem rekomendasi menggunakan metode Collaborative Filtering dapat digunakan untuk berbagai kebutuhan produsen dalam berbagai industri dengan mengambil data interest pengguna dan di proses sedemikian rupa untuk menghasilkan. Diharapkan metode collaborative filtering ini tidak hanya digunakan untuk e-commerce tetapi kedepanya dapat digunakan juga untuk metode penelitian SPK dalam berbagai bidang.

sistem rekomendasi produk dapat dibangun menggunakan metode *item-based collaborative filtering* dengan menerapkan algoritma *K-nearest neighbor*. Algoritma *K-nearest neighbor* ini dipakai untuk menentukan berapa jumlah *neighbor* (produk) terdekat yang digunakan

dalam proses prediksi.

Hasil akurasi yang dicapai dalam pengujian dengan metode ini tergolong akurat.

DAFTAR PUSTAKA

- . [1] Islahudin, H. *Sistem Rekomendasi Bahan Ajar untuk Elearning*. Seminar Nasional Aplikasi Teknologi Informasi 2006 (SNATI 2006) - A125-A130. Yogyakarta. 2006
- . [2] Sania, R., Maharani, W., & Prima K, A. Analisis Perbandingan Metode Pearson dan Spearman Correlation pada Recommender System. Konferensi Nasional Sistem dan Informatika 2010 , 99-105. 2010
- . [3] Santoso, L. W ., Noertjahyana, A., & T andarto, R. *Aplikasi E-Commerce dengan Inteligent Agen Software pada Optik Indo*. SNASTI 2010. Surabaya. 2010: 1-8
- . [4] Uyun, S., Fahrurrozi, I., & Mulyanto, A. Item Collaborative Filtering untuk Rekomendasi Pembelian Buku secara Online. *JUSI*. 2011; 1: 63-70.
- . [5] Winarko, E., Wiranto. *Konsep Multicriteria Collaborative Filtering untuk Perbaikan Rekomendasi*. Prosiding Seminar Nasional Informatika (semnasIF 2010) UPN Yogyakarta. Yogyakarta. 2010:

95-101