

Analisa kinerja metode faktorisasi matriks probabilitas pada sistem rekomendasi = Performance analysis of probabilistic matrix factorization method on recommendation system

Umar Ruswandi, author

Deskripsi Lengkap: <https://lib.ui.ac.id/detail?id=20317195&lokasi=lokal>

Abstrak

Tujuan dari penelitian ini adalah menguji akurasi dari metode faktorisasi matriks probabilitas (pmf) pada data rating film MovieLens. Dalam pmf, data ini dapat direpresentasikan dalam bentuk matriks R , yang berukuran $n \times m$, dimana n adalah banyaknya pengguna sedangkan m adalah banyaknya judul film. Keluaran dari model ini adalah berupa dua buah matriks W dan H . Dimana W adalah matriks fitur pengguna sedangkan H adalah matriks fitur film. Akurasi dari model tercermin dari besarnya norm Frobenius ' $R-WH$ ' pada data testing. Matriks W dan H dapat diestimasi dengan menggunakan Teorema Bayes. Berdasarkan Teorema ini, model yang baik adalah model yang memiliki probabilitas posterior maksimum. Dari eksperimen, kondisi tersebut dicapai saat parameter $k=17$ dan $\lambda=0.2$ dengan $RMSE=0.920661$. Pada nilai $RMSE$ tersebut model ini masuk kategori yang baik dalam memprediksi banyaknya genre dan skor kosong dalam matriks R .

<hr>

The purpose of this study is to test the accuracy of the method of probabilistic matrix factorization (PMF) on MovieLens movie rating data. In PMF, this data can be represents by the sparse matrix R , size $n \times m$, where n is the number of users, while m is the number of movie titles. The output of the model is in the form of two matrices W and H . Where W is the matrix of user features, while H is the matrix of films features. The accuracy of the model is reflected in the size of the Frobenius norm ' $R-WH$ ' in the data testing. Matrices W and H can be estimated by using Bayes theorem. Based on this theorem, a good model is a model that has a maximum posterior probability. From these experiments, the condition is achieved when the parameters $k = 17$ and $\lambda = 0.2$ with $RMSE = 0.920661$. In this model, the $RMSE$ values in the category of good in predicting the number of genre and empty scores in the matrix R .