

Evaluation

Evaluation Metrics for Recommendation Systems¶

Evaluation Metrics Helps to find the Accuracy or Success of Recommender Systems. Predictive accuracy metrics, classification accuracy metrics, rank accuracy metrics, and non-accuracy measurements are the four major types of evaluation metrics for recommender systems.

Predictive Accuracy Metrics

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2}$$

Predictive accuracy or rating prediction measures address the subject of how near a recommender's estimated ratings are to genuine user ratings. This sort of measure is widely used for evaluating non-binary ratings. It is best suited for usage scenarios in which accurate prediction of ratings for all products is critical. Mean Absolute Error (MAE), Mean Squared Error (MSE), Root Mean Squared Error (RMSE), and Normalized Mean Absolute Error (NMAE) are the most important measures for this purpose. In comparison to the MAE metric, MSE and RMSE employ squared deviations and consequently emphasize bigger errors. The error is described by MAE and RMSE in the same units as the obtained data, whereas MSE produces squared units. To make results comparable among recommenders with different rating scales, NMAE normalizes the MAE measure to the range of the appropriate rating scale. In the Netflix competition, the RMSE measure was utilized to determine the improvement in comparison to the Cinematch algorithm, as well as the prize winner.

Classification Accuracy Metrics

		Predicted		
		0	1	
Actual	0	TN	FP Type I error	Specificity = $TN/(TN+FP)$
	1	FN Type II error	TP	Recall or Sensitivity = $TP/(TP+FN)$
		Negative Rate = $TN/(FN+TN)$	Precision = $TP/(TP+FP)$	
Accuracy = $\frac{TP+TN}{TP+FP+TN+FN}$				
F1 - Score = $\frac{2*Recall*Precision}{Recall + Precision}$				

Classification accuracy measures attempt to evaluate a recommendation algorithm's successful decision-making capacity (SDMC). They are useful for user tasks such as identifying nice products since they assess the number of right and wrong classifications as relevant or irrelevant things generated by the recommender system. The exact rating or ranking of objects is ignored by SDMC measures, which simply quantify correct or erroneous classification. This type of measure is particularly well suited to e-commerce systems that attempt to persuade users to take certain actions, such as purchasing products or services.

Rank Accuracy Metrics:

For each user u :

- Generate list of recommendations
- Find rank k_u of its first relevant recommendation (the first rec has rank 1)
- Compute reciprocal rank $\frac{1}{k_u}$

Overall algorithm performance is mean recip. rank:

$$\text{MRR}(O, U) = \frac{1}{|U|} \sum_{u \in U} \frac{1}{k_u}$$

In statistics, a rank accuracy or ranking prediction metric assesses a recommender's ability to estimate the correct order of items based on the user's preferences, which is known as rank correlation measurement. As a result, if the user is given a long, sorted list of goods that are recommended to him, this type of measure is most appropriate. The relative ordering of preference values is used in a rank prediction metric, which is independent of the exact values assessed by a recommender. A recommender that consistently overestimates item ratings to be lower than genuine user preferences, for example, might still get a perfect score as long as the ranking is correct.