

Rating Normalization

Two of the most popular rating normalization schemes that have been proposed to convert individual ratings to a more universal scale are mean-centering and Z-score.

Mean-centering

The idea of mean-centering is to determine whether a rating is positive or negative by comparing it to the mean rating. A raw rating r_{ui} is transformation to a mean-centered one $h(r_{ui})$ by subtracting to r_{ui} the average \bar{r}_u of the ratings given by user u to the items:

$$h(r_{ui}) = r_{ui} - \bar{r}_u$$

Using this approach the user-based prediction of a rating r_{ui} is obtained as

$$\hat{r}_{ui} = \bar{r}_u + \frac{\sum w_{uv} (r_{vi} - \bar{r}_v)}{\sum |w_{uv}|}$$

Z-score normalization

Consider, two users A and B that both have an average rating of 3. Moreover, suppose that the ratings of A alternate between 1 and 5, while those of B are always 3. A rating of 5 given to an item by B is more exceptional than the same rating given by A , and, thus, reflects a greater appreciation for this item. While mean-centering removes the offsets caused by the different perceptions of an average rating, Zscore normalization also considers the spread in the individual rating scales.

In user-based methods, the normalization of a rating r_{ui} divides the user-mean-centered rating by the standard deviation σ_u of the ratings given by user u :

$$h(r_{ui}) = \frac{r_{ui} - \bar{r}_u}{\sigma_u}$$

A user-based prediction of rating r_{ui} using this normalization approach would therefore be obtained as

$$\hat{r}_{ui} = \bar{r}_u + \sigma_u \frac{\sum w_{uv} (r_{vi} - \bar{r}_v) / \sigma_v}{\sum |w_{uv}|}$$