User Modeling and Personalization
8: User Evaluation
Correlation Metrics Exercise

Eelco Herder

L3S Research Center / Leibniz University of Hanover
Hannover, Germany

30 June 2014

Accuracy Metrics - Question

The correlation can be measured using different correlation metrics like *Pearson product-moment correlation coefficient*, *Spearman's rank correlation* or *Kendall tau rank correlation coefficient*.

For which scenarios (datasets, distributions, etc) would you use which metric and why?

Accuracy Metrics - Answers

The most common correlation measure for ratings is the Pearson correlation measure.

Spearman correlations are used if the data does not follow a normal distribution or is otherwise skewed.

Kendalls tau can be used if you have a small data set with a large number of tied ranks.
Kendalls tau is less popular than the Spearmans coefficient, but generally seen as more accurate.

Correlation

Given is the output of one recommender algorithm and the ground truth. Calculate the Pearson correlation, the Spearman's rank correlation and the Kendall's tau coefficient between the ground truth and the recommender output. Assume that higher values represent a higher rank.

Movie	Ground truth	Rec. A
A	2.7	3.4
B	2.8	3.1
C	4.3	7.3
D	5.2	5.1
E	8.0	7.1

Pearson correlation

The most common parametric correlation measure for ratings is the Pearson correlation measure:

$$r(x,y) = \frac{\sum_1^n (x-\overline{x})(y-\overline{y})}{n * \sigma(x)\sigma(y)}$$

The standard deviation σ is given by

$$\sigma = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(x_i-\overline{x})^2}$$

Pearson Correlation - 1

$$
\begin{aligned}
\overline{groundTruth} &= \frac{\sum_1^n groundTruth}{n} = \frac{2.7+2.8+4.3+5.2+8.0}{5} \\
&= 4.6 \\
\overline{recA} &= 5.2 \\
\sigma(groundTruth) &= \sqrt{\frac{\sum_1^n (groundTruth - \overline{groundTruth})^2}{n}} \\
&= \sqrt{\frac{(-1.9)^2 + (-1.8)^2 + (-0.3)^2 + 0.6^2 + 3.4^2}{5}} \\
&\approx 1.942 \\
\sigma(recA) &\approx 1.771
\end{aligned}
$$

Pearson Correlation - 2

$$
\begin{aligned}
r(groundTruth, recA) &= \frac{\sum_1^n (pred - \overline{pred})(act - \overline{act})}{n * \sigma(pred)\sigma(act)} \\
&= \frac{(3.4-5.2)(2.7-4.6)+(3.1-5.2)(2.8-4.6)+(7.3-5.2)(4.3-4.6)+(5.1-5.}{5 * 1.942 * 1.771} \\
&= \frac{3.42 + 3.78 - 0.63 - 0.06 + 6.46}{17.197} \\
&\approx 0.79
\end{aligned}
$$

Spearman's rank correlation

The non-parametric **Spearman's rank correlation coefficient** ρ is used if the data does not follow a normal distribution. The definition is similar to the Pearson correlation. The only difference is that the original values are transformed into ranks and the correlations are computed on the ranks.

$$\rho(x,y) = \frac{\sum_1^n (rank(x) - \overline{rank(x)})(rank(y) - \overline{rank(y)})}{n * \sigma(rank(x))\sigma(rank(y))}$$

Spearman Correlation - 1

For Spearman's rank correlation and Kendall's tau we need a ranked result list. We therefore replace the scores by ranks 1-5:

Movie	Ground truth	Rec. A
A	1	2
B	2	1
C	3	5
D	4	3
E	5	4

$$\overline{groundTruth} = \overline{recA} = 3$$

$$\sigma(groundTruth) = \sigma(recA) = \sqrt{\frac{10}{5}} \approx 1.414$$

Spearman Correlation - 2

$$
\begin{aligned}
r(groundTruth, recA) &= \frac{\sum_1^n (pred - \overline{pred})(act - \overline{act})}{n * \sigma(pred)\sigma(act)} \\
&= \frac{(2-3)(1-3)+(1-3)(2-3)+(5-3)(3-3)+(3-3)(4-3)+(4-3)(5-3)}{5 * 1.414 * 1.414} \\
&= \frac{2+2+0+0+2}{10} \\
&= 0.6
\end{aligned}
$$

Kendall's tau

Let N be the number of joint observations x and y. Let C be the number of *concordant pairs*, pairs of any two of the N items (x,y) for which yields that both $x_i > x_j$ and $y_i > y_j$ (or $x_i < x_j$ and $y_i < y_j$).

And let D be the number of *discordant pairs*, pairs of items for which the above does not yield. Kendal's tau is defined as the difference between the concordant and discordant pairs, divided by all possible item pairs.

$$\tau = \frac{C - D}{\frac{1}{2}N(N-1)}$$

Correlations - Kendall's Tau - 1

For calculating Kendall's tau, we need to consider the order of any two pairs of the ranked lists. In the example we have 10 pairs in the form of (first,second). We create a table that contains the rankings from the ground truth and the recommender algorithms. If the order of the two rating pairs is the same, the pair is *concordant*.

Pair	Concordant	$(GT_{1st}, REC_{1st})(GT_{2nd}, REC_{2nd})$
{A,B}	no	(1,2)(2,1)
{A,C}	yes	(1,3)(2,5)
{A,D}	yes	(1,4)(2,3)
{A, E}	yes	(1,5)(2,4)
{B, C}	yes	(2,3)(1,5)
{B, D}	yes	(2,4)(1,3)
{B, E}	yes	(2,5)(1,4)
{C, D}	no	(3,4)(5,3)
{C, E}	no	(3,5)(5,4)
{D, E}	yes	(4,5)(3,4)

Correlations - Kendall's Tau - 1

For calculating $\tau(groundTruth, recA)$, we count the number of pairs with the same order (concordant pairs), namely $C = |\{\{A,C\},\{A,D\},\{A,E\},\{B,C\},\{B,D\},\{B,E\},\{D,E\}\}| = 7$ and the number of pairs with different order (discordant pairs), namely $D = |\{\{A,B\},\{C,D\},\{C,E\}\}| = 3$.

$$\begin{aligned}\tau(groundTruth, recA) &= \frac{C - D}{\frac{1}{2}N(N-1)} \\ &= \frac{7-3}{0.5 * 5 * (5-1)} \\ &\approx 0.4\end{aligned}$$