

CHANCE PERFORMANCE FOR RANK ACCURACY

ALONA FYSHE

Sometimes in prediction, we score answers by ranking a list of candidate answers by goodness of fit to the predicted answer. For example, in spelling error correction you might be given a misspelled word and then want to rank the list of candidate words so that the correctly spelled word is at the top of the list (or close to it). Under this scenario we may measure the performance of a system with rank accuracy.

Let's say there are N examples to rank, and one prediction for each example. How do we know if our system is performing above chance? If we had built a totally random system, then each time we ranked the candidate words, they'd be in random order. The probability of the correct answer showing up in any of the N spots in the list is $1/N$, a uniform distribution. The expected value of this uniform distribution is

$$\mu = \frac{N + 1}{2}$$

and it has variance

$$\sigma^2 = \frac{N^2 - 1}{12}$$

Each of the N times that we measure the rank of an item, we are drawing from a distribution with mean and variance as described above. Then, we average together those N values to get an estimate of our system's performance. If our system is performing randomly, the distribution of the performance will have mean μ (since the expected value of a sum is just the sum of the expected values). The variance of a sum of variables is the sum of the variable's variance, in this case the sum of the variance would equal $N\sigma$. Since we are averaging the variables rather than summing them, we also have a $1/N$ term which becomes $1/N^2$, making the variance of chance performance:

$$\sigma_{mean}^2 = \frac{\sigma}{\sqrt{N}}$$

This derivation is similar to the derivation for the standard error of the mean.

We can use the formula for the standard error of the mean to calculate a 95% confidence interval around chance accuracy: $\mu \pm 2\sigma_{mean}$. If our system's accuracy is above the upper limit of that confidence interval, we can say our system is performing above chance with 95% confidence.