Problem Set 3

Problem1. MR&S Exercise 8.10 p. 154. I didn't discuss kappa (a standard measure of the level of agreement between two different judgments) in class, but it's in the book pp. 151-152 (page numbers are in the printed copy; the online version is a little different).

Answer:

A. The table can be summarized as below,

	(Judge 2) Yes	(Judge 2) No	(Judge 2) Total
(Judge 1) Yes	2	4	6
(Judge 1) No	4	2	6
(Judge 1) Total	6	6	12

Observed proportion of the times the judges agreed is

$$P(A) = \frac{(2+2)}{12} = \frac{1}{3}$$

$$P(nonrelevant) = \frac{(6+6)}{(12+12)} = \frac{1}{2}$$

$$P(relevant) = \frac{(6+6)}{(12+12)} = \frac{1}{2}$$

Probability that the two judges agreed by chance

$$P(E) = P(nonrelevant)^2 + P(nonrelevant)^2 = 2 * \frac{1}{4} = \frac{1}{2}$$

The kappa measure

$$Kappa = \frac{P(A) - P(E)}{1 - P(E)} = -0.333$$

B. Since a document is considered relevant only if the two judges agree, document 4 in the return set $\{4, 5, 6, 7, 8\}$ is true positive.

$$P = \frac{t_p}{t_p + f_p} = \frac{1}{5}$$

Documents 3 and 4 are relevant in the whole documents, so $t_p + f_n = 2$.

$$R = \frac{t_p}{t_p + f_n} = \frac{1}{2}$$

$$F_1 = \frac{2PR}{P+R} = \frac{2 * \frac{1}{2} * \frac{1}{5}}{(\frac{1}{2} + \frac{1}{5})} = 0.286$$

C. Since a document is considered relevant if either judge thinks it is relevant, all the documents in the return set $\{4, 5, 6, 7, 8\}$ are true positive.

$$P = \frac{t_p}{t_p + f_p} = \frac{5}{5} = 1$$

Documents $\{3, 4, 5, 6, 7, 8, 9, 10, 11, 12\}$ are relevant in the whole documents, so $t_p + f_n = 10$.

$$R = \frac{t_p}{t_p + f_n} = \frac{5}{10} = 0.5$$

$$F_1 = \frac{2PR}{P + R} = \frac{2 * 1 * \frac{1}{2}}{(1 + \frac{1}{2})} = \frac{2}{3} = 0.667$$

Problem 2. Suppose that you want to evaluate an image search engine (e.g. images.google.com). How would you go about this? How does it differ from evaluating a search engine for text documents? Note: I am NOT asking about how you would implement it. I am asking how you would use human evaluators to measure the quality of the results.

Answer:

There are two ways to evaluate the quality of the results. The first one is based on the text described the images. Comparing the users' query words with the keyword around images, the higher relevance of query words and images keywords, the images are ranked higher. The second one is based on the images content, such as color, shape, texture. Comparing the search result with the source image by shape, color or texture, the more similar of the shape, color and texture, the higher rank the searching result gets.

Problem 3.

Answer:

A. Majority class and number of members of the majority class for the three clusters are: Y, 50 (Cluster A); X, 40 (Cluster B); and X, Y, 20, 20 (Cluster C), so

$$Purity = \frac{50 + 40 + 20}{60 + 50 + 40} = 0.733$$

B. The three clusters contain 60, 50, 40 documents, so the total number of "positives" pairs of documents that are in the same cluster is:

$$TP + FP = \binom{60}{2} + \binom{50}{2} + \binom{40}{2} = 3775$$

$$TP = \binom{10}{2} + \binom{50}{2} + \binom{40}{2} + \binom{10}{2} + \binom{20}{2} + \binom{20}{2} = 2475$$

Thus, $FP = 3775 - 2475 = 1300$.

$$FN = \binom{70}{2} - \binom{10}{2} - \binom{40}{2} - \binom{20}{2} + \binom{80}{2} - \binom{50}{2} - \binom{10}{2} - \binom{20}{2} = 3100$$

$\binom{70}{2} - \binom{10}{2} - \binom{40}{2} - \binom{20}{2}$ means that choose 2 from 70 for X minus the sum of same documents in same cluster. $\binom{10}{2}$ is for cluster A, $\binom{40}{2}$ is for cluster B, and $\binom{20}{2}$ is for cluster C.

$\binom{80}{2} - \binom{50}{2} - \binom{10}{2} - \binom{20}{2}$ means that choose 2 from 80 for Y minus the sum of same documents in same cluster. $\binom{50}{2}$ is for cluster A, $\binom{10}{2}$ is for cluster B, and $\binom{20}{2}$ is for cluster C.

Since $TP + TN + FP + FN = 11175$, so $TN = 11175 - 3100 - 3775 = 4300$.

$$P = \frac{t_p}{t_p + f_p} = \frac{2475}{3775} = 0.656$$

$$R = \frac{t_p}{t_p + f_n} = \frac{2475}{2475 + 3100} = 0.444$$

$$F_{\beta(\beta=1)} = \frac{2PR}{P + R} = \frac{2 * 0.656 * 0.444}{(0.656 + 0.444)} = 0.530$$