



به نام خدا



پردیس دانشکده‌های فنی

دانشگاه تهران

دانشکده‌گان فنی

دانشکده مهندسی برق و کامپیوتر



دانشگاه تهران

درس بازیابی هوشمند اطلاعات

پاسخ سوالات تشریحی تمرین اول



سوالات تشریحی

سوال ۱ – Rocchio relevance feedback

پرس و جو و سندهای بازگردانی شده را به صورت برداری تبدیل میکنیم:

$$D_1 = (0,0,0,1,0,1,2)$$

$$D_2 = (0,0,1,0,1,1,1)$$

$$D_3 = (1,1,1,0,1,0,0)$$

$$D_4 = (1,1,0,1,1,0,0)$$

$$q = (0,0,0,0,1,1,1)$$

حال داریم:

$$\begin{aligned} & (0,0,0,0,1,1,1) + \frac{1}{2} * [(0,0,0,1,0,1,2) + (0,0,1,0,1,1,1)] - \frac{1}{2} * [(1,1,1,0,1,0,0) + (1,1,0,1,1,0,0)] \\ &= (0,0,0,0,1,1,1) + \frac{1}{2} * (0,0,1,1,1,2,3) - \frac{1}{2} * (2,2,1,1,2,0,0) \\ &= (0,0,0,0,1,1,1) + (0,0,0.5,0.5,0.5,1,1.5) - (1,1,0.5,0.5,1,0,0) \\ &= (-1,-1,0,0,0.5,2,2.5) \rightarrow (0,0,0,0,0.5,2,2.5) \end{aligned}$$

به همین روش موارد مختلف را مورد بررسی قرار می‌دهیم. برای وزندهی به صورت کلی داریم که بازخورد مثبت مفیدتر است چرا که نمونه‌های منفی بسیار زیادند و در فضا پراکنده‌اند حتی ممکن است ما را منحرف کنند. بازخوردهای مثبت کمتر و متمرکزترند. مرکز آنها میتواند پرس و جو را هدایت کند فلذا لازم است تا وزن بازخورد مثبت را بیشتر از بازخورد منفی در نظر بگیریم.

سوال ۲ - معیارهای ارزیابی

الف) نسبت تعداد اسناد مرتبط بازیابی شده به تعداد کل اسناد بازیابی شده

$$P_1 = \frac{3}{5}, P_2 = \frac{10}{20}, P_3 = \frac{5}{11}$$

ب) نسبت تعداد اسناد مرتبط بازیابی شده به تعداد کل اسناد مرتبط در پایگاه داده

$$R_1 = \frac{3}{10}, R_2 = \frac{10}{10}, R_3 = \frac{5}{10}$$

ج) دقت محاسبه شده بر روی n سند اول بازیابی شده

$$\begin{aligned} P@4(1) &= \frac{2}{4}, P@4(2) = \frac{2}{4}, P@4(3) = \frac{2}{4} \\ P@7(1) &= \frac{3}{5}, P@7(2) = \frac{3}{7}, P@7(3) = \frac{3}{7} \\ P@12(1) &= \frac{3}{5}, P@12(2) = \frac{6}{12}, P@12(3) = \frac{5}{11} \end{aligned}$$

د) دقت محاسبه شده بر روی R سند اول بازیابی شده، که R تعداد کل اسناد مرتبط در پایگاه داده است

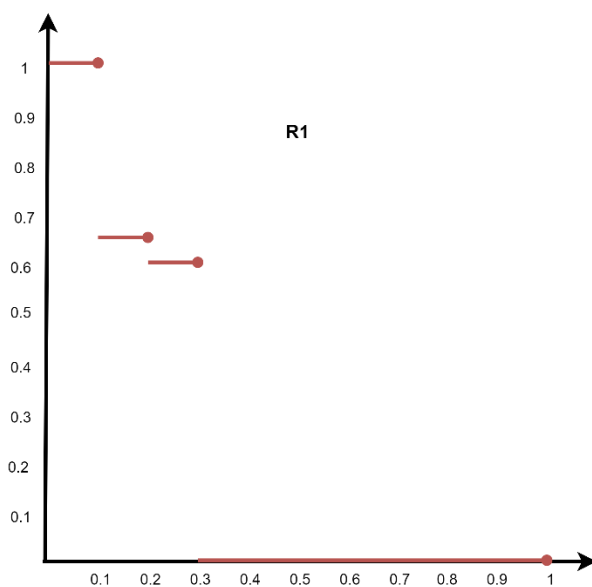
$$RPrec_1 = \frac{3}{5}, RPrec_2 = \frac{5}{10}, RPrec_3 = \frac{5}{10}$$

ه) میانگین دقت همگن بر روی تمام پرسوجوها، که دقت همگن برای هر پرسوجو برابر با میانگین دقت‌های محاسبه شده بر روی همه نقاط بازخوانی است

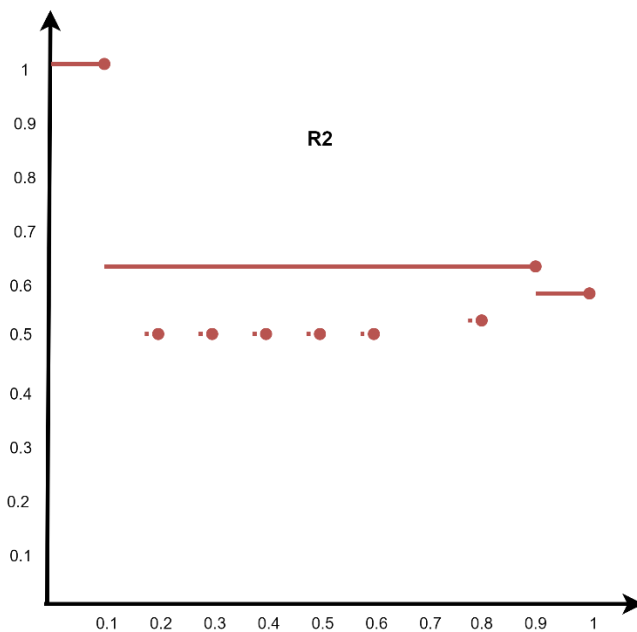
$$\begin{aligned} MAP_1 &= \frac{1 + \frac{2}{3} + \frac{3}{5}}{10} = 0.23 \\ MAP_2 &= \frac{1 + \frac{2}{4} + \frac{3}{6} + \frac{4}{8} + \frac{5}{10} + \frac{6}{12} + \frac{7}{13} + \frac{8}{15} + \frac{9}{16} + \frac{10}{17}}{10} = 0.57 \\ MAP_3 &= \frac{1 + \frac{2}{4} + \frac{3}{5} + \frac{4}{8} + \frac{5}{10}}{10} = 0.31 \\ MAP &= \frac{0.23 + 0.57 + 0.31}{3} = 0.37 \end{aligned}$$

ز) $gMap$ با AP برابر است.

(ج)



نمودار ۱



نمودار ۲

