DR & DNR

SMART implementation →

$$\vec{q}_{opt} = \alpha \vec{q}_0 + \left[\beta \mu(DR) - \gamma \mu(DNR)\right]$$

$\vec{q}_{opt}$ = $\vec{q}_m$ ← modified query.

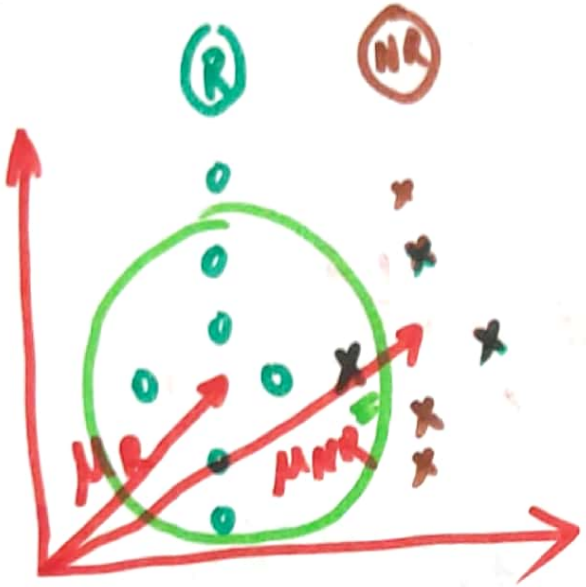
Design choice?

α, β, γ.

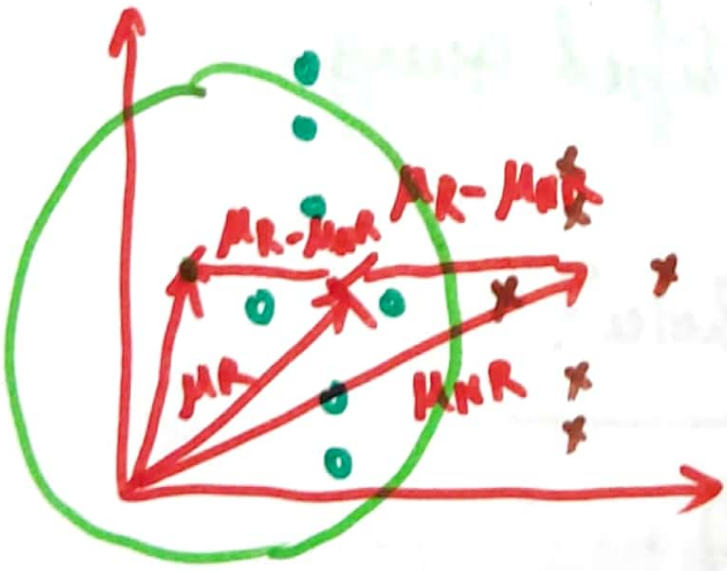
Tradeoff:
(α) (β, γ).

judged documents
more → (β, γ)
less (α).

DR & DNR } wrt q_0 $\mu(DR) + [\mu(DR) - \mu(DNR)]$.



ϵ - neighborhood based on some similarity threshold.



Initial query q_0 ← DR & DNR with respect to q_0

Pseudo relevance feedback.

→ Initial query q_0

→ Get search results from the SE

→ Top ranked K results ↳ DR.

→ Next K results (or even lower) → DNR.

→ Compute Rocchio to get $\vec{q}_m$

→ compare results from $\vec{q}_0$ vs $\vec{q}_m$

$\propto q_0 + \beta MR + \gamma UNR$

-ve entries $\rightarrow$ 0

Most common choice: $\beta = 0.75$, $\gamma = 0.25$

Not very rare: $\beta = 1$, $\gamma = 0$.

- Users are reluctant to give judgements to many queries. (→ judgements may be biased → multiple judgements — more reluctance!)
- q_m computation ← costly
  → recalling q_m ← expensive.

human expertise → time consuming & expensive.

automatic.

↓

construction.

HOSPITAL → HOSPITAL, MEDICAL, EMERGENCY, NURSE, DOCTOR.

INTEREST RATE → INTEREST RATE, FASCINATED, DATING, LOVELIFE

Query Expansion.

↳ global Knowledge.

Explicit (near)-synonyms.

$q_0 \rightarrow w_0 \ldots w_m$.

↑ expanded to include near-synonyms words from the dictionary of near synonyms words.

thesaurus.

H_2 → two words are similar if they bear similar grammatical relationship with same set of words.

"apple" ≅ "pear"

peel, eat, harvest, trade → apples & pears.

$$PMI(w_1, w_2) = -\log_2 \frac{P_{corpus}(w_1, w_2)}{P_{corpus}(w_1)\, P_{corpus}(w_2)}$$

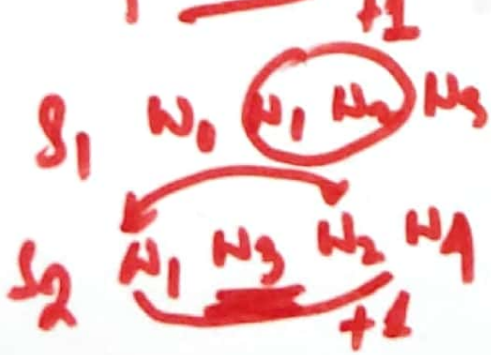
↓

pointwise mutual information.

$$P_{corpus}(w_1, w_2) = \frac{freq(w_1, w_2)}{N}$$

$$P_{corpus}(w_1) = \frac{freq(w_1)}{N}$$

$$P_{corpus}(w_2) = \frac{freq(w_2)}{N}$$



H1: Co-occurrence.

H2: Grammatical relationship

H1: Two words are similar if they have similar context, i.e., the neighboring words of these two words are similar.

"car" ≃ "motorcycle"

"road", "gas", "license" etc.

herb

↳ herbal remedies.