

CS 533 Information Retrieval

The Use of MMR,
Diversity-based Reranking
for Reordering Documents
and Producing Summaries

by Jaime Carbonell & Jade Goldstein

Devrim Şahin, Selçuk Emre Solmaz

Relevance vs. Diversity

Relevance: Results are similar to the query

Diversity: Results are dissimilar from each other

Diversity reduces duplicative information in the result set, but decreases accuracy: **TRADEOFF**

HOW DIVERSE?

We need user-tunable diversity: **MMR**

Suggested by Jaime Carbonell & Jade Goldstein

Maximal Marginal Relevance (MMR)

$$\text{MMR} = \arg \max_{D_i \in R \setminus S} \left[\lambda \text{Sim}_1(D_i, Q) - (1 - \lambda) \max_{D_j \in S} \text{Sim}_2(D_i, D_j) \right]$$

D_i : Documents in the collection, Q : Query,

R : Relevant documents, S : Current result set

Greedy, user-tunable

High lambda = Higher accuracy

Low lambda = Higher diversity

Document Reordering

“Experiment”

Participants: 5 undergraduate students

4 of them preferred MMR over relevance-only search

3 of them “clearly discovered the differential utility of diversity search and relevance-only search”

Document Summarization

Segment the document into passages (sentences)

Use MMR + cosine similarity

User generated query

Experiment:

May 1998 SUMMAC Conference

Government-run evaluation of 15 summarization systems

Highest-utility query-relevant summaries (F-score of 0.73)

Works better for longer documents

Demo

Thank you for listening!

