

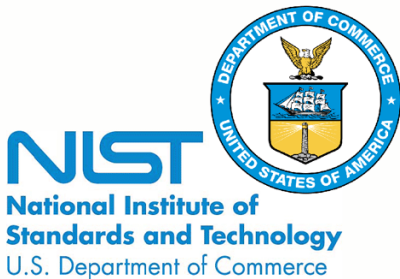


Information Retrieval

Natural Language Processing: Jordan
Boyd-Graber
University of Maryland
EVALUATION

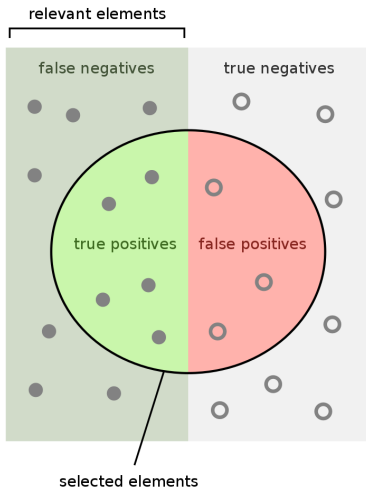
Slides adapted from Jimmy Lin

Evaluating Results: Intrinsic



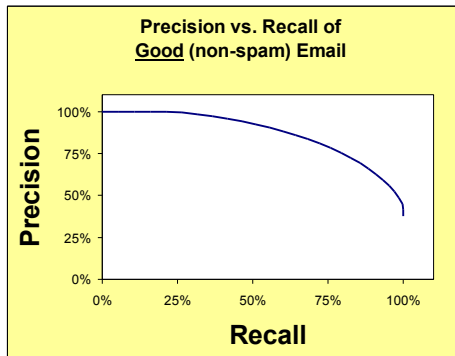
- TREC: set of “gold” relevant documents
- How many of the documents found?
- Annual bake-off of IR systems

Relevance Terminology



- TP: True Positive
- FP: False Positive
- TN: True Negative
- FN: False Negative

Precision and Recall



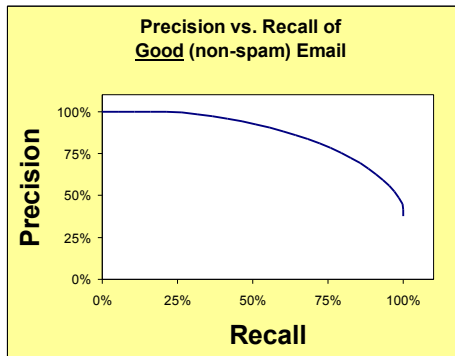
- Precision: Of what you returned, how much was right?

$$P = \frac{|TP|}{|TP| + |FP|} \quad (1)$$

- Recall: Of what could be right, how much did you find?

$$P = \frac{|TP|}{|TP| + |FN|} \quad (2)$$

Precision and Recall



- Precision: Of what you returned, how much was right?

$$P = \frac{|TP|}{|TP| + |FP|} \quad (1)$$

- Recall: Of what could be right, how much did you find?

$$P = \frac{|TP|}{|TP| + |FN|} \quad (2)$$

- *F*-measure: geometric mean

Evaluating Results: Extrinsic

- IR is often used to find answers to questions
- But it takes a human to read the results
 - If you know what answer is, you can search for similar questions
 - Build machines to read the answer
- We'll see both later in the class