EVALUATION IN IR Introduction to Information Retrieval CS 150 Donald J. Patterson

Content adapted from Hinrich Schütze <u>http://www.informationretrieval.org</u>

Unranked retrieval - ROC curve Receiver Operating Characteristic (ROC) curve



Unranked retrieval - ROC curve

Receiver Operating Characteristic (ROC) curve



Ranked Retrieval

- Precision and Recall are set-based measures
 - They are computed independent of order
 - But, web search return things in lists
 - Lists have order.
 - A better metric of user happiness/relevance is warranted

Ranked Retrieval

- Let's use our existing metrics and extend them to ranked retrieval
 - In one system we can get many samples
 - We can get the top X results:
 - X= 10, 20, 30, 40, etc...
 - Each one of those sets has a precision and recall value
 - Each of those sets corresponds to a point on the ROC curve.



Ranked Retrieval

• Each of those sets corresponds to a point on the ROC



Ranked Retrieval

- One option is to average the precision scores at discrete points on the ROC curve
- But which points?
- We want to evaluate the system, not the corpus
- So it can't be based on number_{0%}

of documents returned



Ranked Retrieval - 11 point precision

- Evaluate based on precision at defined recall points
- Average the precision at 11 points
- This can be compared across corpora
 - because it isn't based on corpus size or number of

results returned



Ranked Retrieval - Mean Average Precision

- Why just 11 points?
- Why not average over all points?
- This is roughly equivalent to measuring the area under the

curve.



Precision

Less junk



Ranked Retrieval - Precision at k

- Users don't care about results past a page or two
- So area under the curve is too naive.
- Let's evaluate precision with k results
 Recall
 instead.
- Highly dependent on number of relevant documents
- If k is 20 and relevant docs is 8

best score is 8/(8+12) = 0.4



Precision

Ranked Retrieval - Precision at R

- We know the number of relevant documents, r, so
- rather than looking at k results let's look at the top r results
- If r is 20
 - best score is 20/(20) =1.0
 - best score is always 1.0



Precision

Ranked Retrieval - Precision at R

- It turns out that Precision at R is the break-even point
 - When Precision and Recall are equal
 - Do we care about this point for any rational reason?



Precision

Critiques of relevance

- Is the relevance of one document independent of another?
- Is a gold standard possible?
 - Is a gold standard static?
 - Uniform?
 - Binary?
- Perhaps relevance as a ranking is better.
- Relevance versus marginal relevance

what does another document add?

Refining a deployed system

- Once you have a system, with metrics, how do you consider changing the system to improve the metrics?
- A common approach is A/B testing.
 - This is done by Google for clients and Amazon for itself and probably many others.
- The idea:
 - Treat a small number of your users as experiments.
 - Have them use the different system.
 - Evaluate metrics on experimental group.







17.

Fulfillment by Amazon

WebStore by Amazon

\$5 or \$10 Bonus on Southern Living

Subscribe to Southern Living and the second second later where the second se

Evaluation in IR Google donald.j.patterson.iii@gmail.com | Help | Contact Us | Sign Out Customer ID: 475-906-9660 Campaign Management My Account Reports Analytics Campaign Summary | Tools | Conversion Tracking | Website Optimizer Search Tools > Website Optimizer Website Optimizer (beta): Experiment List Common Questions ✓ **Congratulations!** You've signed up for Website Optimizer and are ready to create your experiment. As you move through the steps, your information will be saved. How do I create an experiment? + Create a new experiment Do I need my webmaster's help? View: All experiments • Delete [?] Helpful Links Finish Conv. Page Conv. Experiment Status Visitors Rate Time Website Optimizer Help No experiments Website Optimizer Demo (Flash - English Only) Click Create experiment to get started. Website Optimizer User Guide My Client Center User Access [?] (Flash - English Only) Tell us about Website Optimizer. Quick Start Guide Take a survey Installation Guide Testing Guides and Strategies Discussion Forum Website Optimizer Home Page Send Us Feedback ©2007 Google - AdWords Home - Editorial Guidelines - Privacy Policy - Contact Us

Snippets

• Little bits of text that summarize the page

Informatics - Wikipedia, the free encyclopedia Informatics includes the science of information, the practice of information ... Informatics studies the structure, behavior, and interactions of natural and ... en.wikipedia.org/wiki/Informatics - 35k - Cached

- They function as an implicit tool for users to rank the results on their own (among those visible)
- The user does the final ranking
- Users are still biased by presented order though.

Snippets

- The goal of snippet generation is
 - present the most informative bit of a document in light of the query
 - present something which is self-contained
 - i.e., a clause or a sentence
 - present something short enough to fit in output
 - be fast, accurate (where are the snippets stored?)
- Challenges
 - Multiple occurrences of keyword in document
 Poor English or other language grammar

Snippets

- Snippets can be static
 - A snippet for a web page is precompiled and always the same.
- Snippets can be dynamic
 - Depends on the query
 - "informatics"

Informatics - Wikipedia, the free encyclopedia

Informatics includes the science of information, the practice of information ... Informatics studies the structure, behavior, and interactions of natural and ... en.wikipedia.org/wiki/Informatics - 35k - Cached

"informatics definition"

Informatics - Wikipedia, the free encyclopedia Usage has since modified this definition in three ways. ... broader definition at the launch of its School of Computing and Informatics in September 2006. ... en.wikipedia.org/wiki/Informatics - 35k - Cached

Snippets

- Snippets may contain
 - A few sentences from the web page
 - Meta data about the page
 - Author, Date, Title
 - Output of a text-summarization algorithm
 - Advanced technology that attempts to write snippets
 - Images from the document

