The Maximum Entropy Method for Analyzing Retrieval Measures

Javed Aslam
Emine Yilmaz
Vilgil Pavlu
Northeastern University
Evaluation Measures:
Top Document Relevance
Evaluation Measures:
First Page Relevance
Natural Question...

• Many measures of retrieval performance
  • 23 standard measures used in TREC
• Are some measures “better” than others?
  • system-oriented vs. user-oriented measures
• Q: What can be learned from a measure?
• A: Good overall measures:
  • reduce uncertainty about underlying phenomenon
  • allow one to infer underlying phenomenon
• E.g., Health:
  • BMI vs. blood pressure vs. cholesterol vs. shoe size
Research Goals

• What can be reasonably inferred from a measure?
  • maximum entropy method...

• How good are those inferences?
  • compare inferences to reality (e.g., TREC)

• Assess quality of measures
  • error, prediction, reduction in uncertainty
Outline

• Introduction
• Standard measures for query retrieval
• The maximum entropy method
  • dice example
  • measures as constraints
• MEM for query retrieval measures
• Experimental results
Evaluation Measures Setup

- Ranked list of retrieved documents
- Binary relevance judgments
- Good performance:
  - “many” relevant docs “high” in list
Traditional IR Measures

- Precision of top k documents, for k:
  - 5, 10, 15, 20, 30, 100, 200, 500, 1000

- R-precision:
  - precision of top R documents, where $R = \#$ relevant docs

- Average precision:
  - average of precisions at all R relevant documents...
Traditional IR Measures: Average Precision

List:

\[ AP = \frac{1 + \frac{2}{3} + \frac{3}{6} + \frac{4}{10}}{4} \approx 0.6417 \]
Most Commonly Used IR Measures

• Average precision
• Precision at 10 (30, etc.) documents
• R-precision
• Precision-recall curves...
Visualizing Retrieval Performance: Precision-Recall Curves
Analyzing Retrieval Measures: Setup

• A list or its P-R curve defines performance

• How much does a measure reduce one’s uncertainty in the underlying list or its P-R curve?

• Good measures: large reduction in uncertainty

• Poor measures: little or no reduction in uncertainty

• How to measure reduction in uncertainty?

• Maximum entropy method...
Outline

• Introduction
• Standard measures for query retrieval
• The maximum entropy method
  • dice example
  • measures as constraints
• MEM for query retrieval measures
• Experimental results
Maximum Entropy Method: Dice Example

- Given an unknown six-sided dice, what is probability for each die face (1, 2, 3, 4, 5, 6)?
- Under-constrained problem
  - most "reasonable" answer is uniform (1/6, 1/6, ..., 1/6)
- What if average die roll is 4.5?
  - problem still under-constrained, but what is the most "reasonable" answer?
  - maximum entropy method to the rescue...
Maximum Entropy Method

- Goal: infer probability distribution (belief) from statistics (measures or constraints) over that distribution
- Uses: prediction, coding, gambling, etc.
- MEM dictates the most “reasonable” solution
Back to Dice Example

- Average die roll is 4.5; what is distribution?
- One solution:
  - Principle of “maximal ignorance:” pick distribution which is least predictable (most random) subject to constraints
  - How to measure randomness? Entropy \( H(\vec{p}) = \sum_{i=1}^{6} p_i \lg(1/p_i) \)
  - Thus, max entropy distribution subject to constraints
Mathematical Justification

• Entropy Concentration Theorems
  • “weak” and “strong”
• Nature favors maximum entropy solutions
  • e.g., temperature and particle speed
Max Entropy Distributions: Dice Examples

\[ E[X] = 3.5 \]
\[ E[X] = 4.5 \]
\[ E[X] = 5.5 \]
Outline

• Introduction
• Standard measures for query retrieval
• The maximum entropy method
  • dice example
  • measures as constraints
  • properties of the maximum entropy distribution
• MEM for query retrieval measures
• Experimental results
MEM for IR Measures: An Analogy

<table>
<thead>
<tr>
<th>Problem</th>
<th>Events</th>
<th>Distribution</th>
<th>Constraint</th>
</tr>
</thead>
<tbody>
<tr>
<td>Dice</td>
<td>die faces</td>
<td>over die faces</td>
<td>expected die roll</td>
</tr>
<tr>
<td>IR</td>
<td>lists</td>
<td>over lists</td>
<td>expected AP, RP, P@10</td>
</tr>
</tbody>
</table>
IR Setup: Distribution over Lists

- possible relevance list
- distribution over lists
- independence assumption
  - \( p(x_1, x_2, \ldots, x_m) = p_1(x_1) \cdot p_2(x_2) \cdots p_m(x_m) \)
- probability-at-rank
  - \( p_i = p_i(x_i) \)
How to Find Max Ent Dist?

• Assumption: \( p(x_1, x_2, \ldots, x_m) = p_1(x_1) \cdot p_2(x_2) \cdots p_m(x_m) \)

• Entropy: \( H(p(x_1, \ldots, x_m)) = \sum_{i=1}^{m} H(p_i) \)

• Constraints:
  • measure (AP, RP, P@10)
  • total number of relevant documents \( R \)
Setup for PC@10 Constraint

- Maximize: \( \sum_{i=1}^{m} H(p_i) \)

- Subject to:
  - \( \sum_{i=1}^{10} p_i = 10 \cdot PC(10) \)
  - \( \sum_{i=1}^{m} p_i = R \)
Setup for RP Constraint

• Maximize: \[ \sum_{i=1}^{m} H(p_i) \]

• Subject to:

\[ \sum_{i=1}^{R} p_i = R \cdot RP \]

\[ \sum_{i=1}^{m} p_i = R \]
Setup for AP Constraint

• Maximize: \( \sum_{i=1}^{m} H(p_i) \)

• Subject to:
  \[
  \sum_{i=1}^{m} \left( \frac{p_i}{i} \left( 1 + \sum_{j=1}^{i-1} p_j \right) \right) = R \cdot \text{AP}
  \]

• \( \sum_{i=1}^{m} p_i = R \)
Solutions

• Analytical: Lagrange multipliers
• Numerical: MatLab, Mathematica, etc.
Outline

- Introduction
- Standard measures for query retrieval
- The maximum entropy method
  - dice example
  - measures as constraints
  - properties of the maximum entropy distribution
- MEM for query retrieval measures
- Experimental results
Experimental Results

- Take actual lists from TREC conference
- Compute AP, RP, P@10
- Find max ent dist for these constraints
- Compare to actual list
Inferred Probability at Rank

Probability at rank curve for system fub99a Query 435
Inferred Probability at Rank
Actual and Inferred P-R Curves

Precision at recall curve for system INQ601 Query 445

- Actual dist
- AP maxent dist
- RP maxent dist
- PC maxent dist
P-R Curves

Precision at recall curve for system fub99a Query 435

- actual dist
- AP maxent dist
- RP maxent dist
- PC maxent dist
Precision at recall curve for system MITSLStd Query 404

- actual dist
- AP maxent dist
- RP maxent dist
- PC maxent dist

P-R Curves
P-R Curves

Precision at recall curve for system plt8ah1 Query 446

- actual dist
- AP maxent dist
- RP maxent dist
- PC maxent dist
Future Work & Questions?