CMSC828E
Ranking in Information Retrieval

Jian Li
University of Maryland, College Park
Sep. 2010
Typical IR Problem

Document representation

Query representation

Query

Document collection

Query Answer
Outline

• The probability ranking principle
• The binary independence model
• When PRP fails?
  • Multiple (or uncertain) intents
  • Correlated documents
• Attempt to remedy
  • Portfolio theory
  • Chen-Karger Model
  • Multi-arm bandits Model
Probability Ranking Principle

- Collection of Documents
- User issues a query
- A Set of documents needs to be returned
- **Question:** In what order to present documents to user?

Introduction to Information Retrieval.
Manning, Raghavan and Schutze. Chapter 11.
Probability Ranking Principle

- **Question:** In what order to present documents to user?
- Intuitively, want the “best” document to be first, second best - second, etc...
- Need a formal way to judge the “goodness” of documents w.r.t. queries.
- **Idea:** Probability of relevance of the document w.r.t. query
Probability Ranking Principle

- **PRP**: rank documents by their estimated probability of relevance ($Pr(R=1 \mid d, q)$)
  - $d$: document, $q$: query

- **THM**: The PRP is optimal, in the sense that it minimizes the expected 0/1 loss (also known as Bayes risk).
  - 0/1 loss: In the top-$k$ result for any fixed $k$,
    - number of nonrelevant docs retrieved + number of relevant docs left out

- If we think the 0/1 loss in a possible world as the distance function between the retrieved answer and the true answer, ranking according to PRP is equivalent to the consensus answer.
Outline

• The probability ranking principle
• The binary independence model
• When PRP fails?
  • Multiple (or uncertain) intents
  • Correlated documents
• Attempt to remedy
  • Portfolio theory
  • Chen-Karger Model
  • Multi-arm bandits Model
The Binary Independence Model

- How to estimate $P(R=1|d, q)$
- Bayesian probability formulas

\[ p(a | b)p(b) = p(a \cap b) = p(b | a)p(a) \]

\[ p(a | b) = \frac{p(b | a)p(a)}{p(b)} \]

- Odds: \( O(y) = \frac{p(y)}{p(\overline{y})} = \frac{p(y)}{1 - p(y)} \)
Binary Independence Model

- Traditionally used in conjunction with PRP
- "Binary" = Boolean: documents are represented as binary vectors of terms:
  - \( \tilde{x} = (x_1, \ldots, x_n) \)
  - \( x_i = 1 \text{ iff term } i \text{ is present in document } x \).

- "Independence": terms occur in documents independently
- Different documents can be modeled as same vector.
Binary Independence Model

- Queries: binary vectors of terms
- Given query $q$,
  - for each document $d$ need to compute $p(R|q,d)$.
  - replace with computing $p(R|q,x)$ where $x$ is vector representing $d$
- Interested only in ranking
- Will use odds:

$$O(R|q,\bar{x}) = \frac{p(R|q,\bar{x})}{p(NR|q,\bar{x})} = \frac{p(R|q)}{p(NR|q)} \cdot \frac{p(\bar{x}|R,q)}{p(\bar{x}|NR,q)}$$
Binary Independence Model

\[ O(R \mid q, \bar{x}) = \frac{p(R \mid q, \bar{x})}{p(NR \mid q, \bar{x})} = \frac{p(R \mid q)}{p(NR \mid q)} \cdot \frac{p(\bar{x} \mid R, q)}{p(\bar{x} \mid NR, q)} \]

- Using Independence Assumption:

\[ \frac{p(\bar{x} \mid R, q)}{p(\bar{x} \mid NR, q)} = \prod_{i=1}^{n} \frac{p(x_i \mid R, q)}{p(x_i \mid NR, q)} \]

- So:

\[ O(R \mid q, d) = O(R \mid q) \cdot \prod_{i=1}^{n} \frac{p(x_i \mid R, q)}{p(x_i \mid NR, q)} \]

Constant for each query

Needs estimation
**Binary Independence Model**

\[ p_i = p(x_i = 1 \mid R, q); \quad r_i = p(x_i = 1 \mid NR, q); \]

Under some simplifying assumptions, we can get

\[ O(R \mid q, \bar{x}) = O(R \mid q) \cdot \prod_{x_i=q_i=1} \frac{p_i(1-r_i)}{r_i(1-p_i)} \cdot \prod_{q_i=1} \frac{1-p_i}{1-r_i} \]

- Only quantity to be estimated for rankings
- Constant for each query
- Retrieval Status Value:

\[ RSV = \log \prod_{x_i=q_i=1} \frac{p_i(1-r_i)}{r_i(1-p_i)} = \sum_{x_i=q_i=1} \log \frac{p_i(1-r_i)}{r_i(1-p_i)} \]

See the book for the derivation.
Binary Independence Model

\[ p_i = p(x_i = 1 | R, q); \quad r_i = p(x_i = 1 | NR, q); \]

• For each term \( i \) look at the following table:

<table>
<thead>
<tr>
<th>Documents</th>
<th>Relevant</th>
<th>Non-Relevant</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>( X_i=1 )</td>
<td>( s )</td>
<td>( n-s )</td>
<td>( n )</td>
</tr>
<tr>
<td>( X_i=0 )</td>
<td>( S-s )</td>
<td>( N-n-S+s )</td>
<td>( N-n )</td>
</tr>
<tr>
<td>Total</td>
<td>( S )</td>
<td>( N-S )</td>
<td>( N )</td>
</tr>
</tbody>
</table>

• Estimates:

\[ p_i \approx \frac{s}{S} \quad r_i \approx \frac{(n-s)}{(N-S)} \]

\[ c_i \approx K(N, n, S, s) = \log \frac{s/(S-s)}{(n-s)/(N-n-S+s)} \]

Add 0.5 to every expression
Outline

- The probability ranking principle
- The binary independence model
- When PRP fails?
  - Multiple (or uncertain) intents
  - Correlated documents
- Attempt to remedy
  - Portfolio theory
  - Chen-Karger Model
  - Multi-arm bandits Model
When PRP fails?

- **Multiple (or uncertain) intents**
  - E.g. Clinton
- **Correlated documents**
- **A counter example (W.S.Cooper)**

Any type 1 user would be satisfied with any doc 1-9

**Type 1: 100 users**

Any type 2 user would be satisfied with doc 10

**Type 2: 50 users**

PRP: d1, d2, d3, ..., d10

Optimal: d1, d10, d2, d3, ...

Diversification

\[ P(R=1) = \frac{2}{3} \]

\[ P(R=1) = \frac{1}{3} \]
When PRP fails

- A more detailed discussion, see When is the probability ranking principle suboptimal. Gordon and Lenk, ’92.
Outline

• The probability ranking principle
• The binary independence model
• When PRP fails?
  • Multiple (or uncertain) intents
  • Correlated documents
• Attempt to remedy
  • Portfolio theory
  • Chen-Karger Model
  • Multi-arm bandits Model
Portfolio Theory

Available securities: Rate of return: random variable!

How to invest your money?
Portfolio Theory

Available securities: Rate of return: random variable!

Assume Xi are independent

X1
P(r=2)=.5
P(r=1)=.5

X2
P(r=2)=.5
P(r=1)=.5

X3
P(r=2)=.5
P(r=1)=.5

X4
P(r=2)=.5
P(r=1)=.5

Two strategy: (assume we have 1$)

(1): Invest 1$ to X1: E[return]=1.5 Var[return]=0.25
(2): Invest .25$ to each Xi E[return]=1.5 Var[return]=0.25/4=0.0625

Budget: $  $

Do not put all your eggs in one basket!

The risk is much smaller
Portfolio Theory

- What to optimize?
  - Maximize $E[R]$
  - Minimize $\text{Var}[R]$
  - Minimize $\text{Var}[R]$, subject to $E[R] \geq t$
  - Maximize $E[R]$, subject to $\text{Var}[R] \leq t$
  - Maximize $E[R] - b \times \text{Var}[R]$
    - $b > 0$: risk averse
    - $b < 0$: risk loving
Portfolio Theory

Efficient frontier

Attainable E, V combinations

E + bVar = C \ (b > 0)

E + bVar = C \ (b < 0)
Portfolio Theory in IR

Find an assignment.

Find a ranking.

Portfolio Theory of Information Retrieval. Wang and Zhu. SIGIR 09
Portfolio Theory in IR

• The algorithm in [Wang&Zhu]
  • For i=1 to n
    • For position i, greedily choose the document that maximizes the marginal increase of the objective value among remaining documents.

• It is not optimal
  • Open: is the problem NP-hard? Any performance guarantee for the greedy algorithm?
Outline

- The probability ranking principle
- The binary independence model
- When PRP fails?
  - Multiple (or uncertain) intents
  - Correlated documents
- Attempt to remedy
  - Portfolio theory
  - Chen-Karger Model
  - Multi-arm bandits Model
The “less is more” Model

- User may be satisfied with *one* relevant result
  - Navigational queries, question/answering
- We want *diversity* in result set
  - Better to satisfy different users with different interpretations, than one user many times over
- Let us just change our objective:
  - Maximize the prob. of finding $h$ documents among the top-$k$ answer (for small $h$).
  - This naturally introduced diversification into the ranking.
- Check Cooper’s example.

Less is more: probabilistic models for retrieving fewer relevant documents.
Chen and Karger. SIGIR06
Outline

• The probability ranking principle
• The binary independence model
• When PRP fails?
  • Multiple (or uncertain) intents
  • Correlated documents
• Attempt to remedy
  • Portfolio theory
  • Chen-Karger Model
  • Multi-arm bandits Model
The Multi-armed Bandits Model

• The multi-armed bandits problem

K gambling machines. Playing machine $i$ yield rewards $x_{i1}$, $x_{i2}$, $\ldots$ which are i.i.d according to some unknown distribution.
Find a strategy (how to play) to maximize the expected payoff.

There are strategies that can achieve:

$$E[\text{payoff up to time } T] \geq \text{OPT} - R(T)$$

Where $R(T) = o(T)$. 
The Multi-armed Bandit Model

- A set of documents $D=\{d_1, \ldots, d_n\}$.
- A population of users.
- Users of type $i$ will click $d_j$ with prob $p_{ij}$
- Each time the system presents to the user a top-$k$ list.
- The user click the first doc she likes.
- Objective: Maximizes $E[\#\text{users who click at least once}]$

Learning diverse ranking with multi-armed bandits.
Radlinski, Kleinberg and Joachims. ICML08.
The Multi-armed Bandit Model

- For each time slot, we run an MAB instance.

Top-k slots:

- Each doc corresponds to an arm.
- If the user click any doc in the list, we get payoff 1.
- There exists a strategy that achieves an expected payoff of at least \((1-1/e)\text{OPT} - O(k \times \sqrt{T \log n})\)
Notes

- Parts of this slides use material from Alexander Dekhtyar’s slides on probabilistic information retrieval.