11-442 / 11-642 / 11-742:
Search Engines

Diversity

Jamie Callan and Chenyan Xiong
Carnegie Mellon University
callan@cs.cmu.edu

1

Topic Outline

- Introduction
- Diversity Evaluation Metrics
- Implicit Methods
  - – Maximum Marginal Relevance (MMR)
  - – Learning to Rank for diversification
- **Explicit Methods**
  - – Query intents (subtopics) discovery
  - – xQuAD
  - – PM-2

2

2

Explicit Query Intents

Some diversification methods assume that query intents are known

- Query intents are explicit

How can query intents be known?

- Analysis of search logs (covered in a few weeks)

How can this problem be studied without search logs?

- Query intents provided by TREC (which got them from Bing)
- Query intents inferred from commercial search engine suggestions

Explicit Query Intents: The TREC Web Track Diversity Task

The 2009-2014 TREC Web Tracks had a diversity task

- Intended to encourage research on diversification

Each year, 50 information needs were created

- Most information needs had multiple intents
  - Ambiguous: Unrelated interpretations of the query
  - Faceted: Related interpretations of the query
- Information needs with multiple intents had two types of intents
  - Navigational
  - Informational

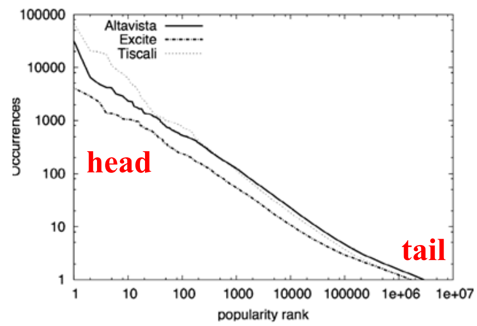
Explicit Query Intents: TREC Information Needs

Queries were manually chosen from Bing's search log

- Picked from median queries (by frequency)
  - Head queries are too easy
  - Tail queries are too rare

Subtopic generation process

- Cluster Bing's query log,
- Manually pick queries from top clusters



5

5

Explicit Query Intents: TREC Information Needs

Topic 15, type=ambiguous

nav: **navigational**
inf: **informational**

- **Query:** espn sports
- **Description:** I'm looking for various sports scores and information from the ESPN Sports site.
- **Subtopic, type=nav:** The ESPN Sports home page.
- **Subtopic, type=inf:** College football and basketball scores.
- **Subtopic, type=inf:** NBA basketball standings.
- **Subtopic, type=inf:** Baseball scores and information on upcoming live broadcast games.
- **Subtopic, type=inf:** Information on NASCAR races.
- **Subtopic, type=inf:** Fantasy football leagues.

(http://trec.nist.gov/data/web/09/wt09.topics.full.xml)

6

6

Explicit Query Intents: TREC Information Needs

nav: navigational
inf: informational

Topic 1, type=faceted

- **Query:** obama family tree
- **Description:** Find information on President Barack Obama's family history, including genealogy, national origins, places and dates of birth, etc.
- **Subtopic, type=nav:** Find the TIME magazine photo essay "Barack Obama's Family Tree".
- **Subtopic, type=inf:** Where did Barack Obama's parents and grandparents come from?
- **Subtopic, type=inf:** Find biographical information on Barack Obama's mother.

(http://trec.nist.gov/data/web/09/wt09.topics.full.xml)

Explicit Query Intents: TREC Information Needs

Note the two types of topics

- **Ambiguous:** Unrelated interpretations of the query
  – E.g., “michael jordan”, “avp”, “espn sports”
- **Faceted:** Related interpretations of the query
  – E.g., “carnegie mellon university”, “arizona game and fish”, “obama family tree”

Different types might require different methods

- TREC hoped to focus attention on this problem
- A difficult problem, mostly ignored by researchers (so far)

Explicit Query Intents: TREC Information Needs

In TREC evaluations, the search engine <u>does not know</u> the multiple intents for an information need

- It sees only the query, e.g., "espn sports"
- <u>The assessor</u> knows the different intents
  - Used to form relevance assessments

What information can <u>your</u> search engine use to diversify the ranking?

Explicit Query Intents: Commercial Search Engines

Google and Bing provide <u>suggested</u> and <u>related</u> queries

- Probably they indicate common search intents

Query suggestions

es|pn
- espn
- espn3
- espn fantasy
- espn mlb

espn s
- espn scores
- espn sports
- espn soccer
- espn schedule

espn sp|orts
- espn sports
- espn sports center
- espn sportsnation
- espn sports science

Related queries

Searches related to espn sports

espn boxing	espn soccer
fox sports	espn sports golf
espn sports live	yahoo sports
espn scores	cbs sports

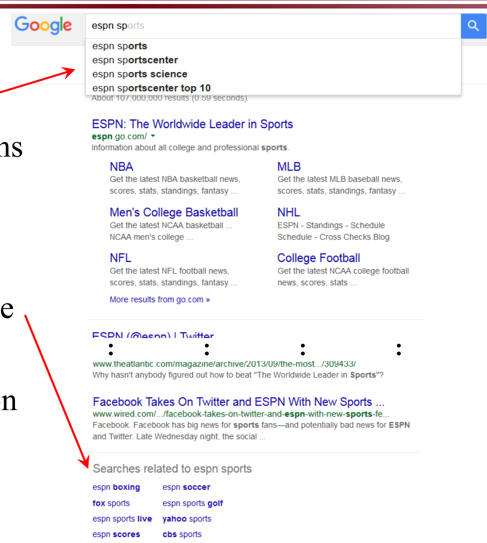
Explicit Query Intents: Commercial Search Engines

Suggested queries

- Appear at the top of the page
- Often prefix-oriented suggestions
  - High precision, low recall

Related queries

- Appear at the bottom of the page
- A wider range of suggestions
  - Higher recall, lower precision



Explicit Query Intents: Commercial Search Engines

Suggested and related queries produced by three web search engines

Search Engine	Query Type	Avg # of Queries	Avg Query Length	Avg # of Matched Pages
A	related	7.4	2.9	10.4 M
	suggested	6.7	3.0	8.6 M
B	related	9.7	2.6	10.3 M
	suggested	9.3	3.3	7.1 M
C	related	15.9	1.8	18.1 M
	suggested	8.9	3.3	9.5 M

- An average of 14-25 short suggested and related queries

(Santos, et al., 2010)

Explicit Query Intents: Automatic Generation

How can query intents be generated without a query log?

- Only large search companies have large query logs
- Even Google doesn't have much information about tail queries

This topic is challenging for several reasons

- Query intents are very personal
  – Different people have different interpretations
- There are <u>many</u> possible subtopics
  – Which ones does the gold standard want?
- There are many <u>types</u> of subtopics
  – E.g. ambiguous vs. facets vs. tasks
- Which ones does the gold standard want?

13

13

Explicit Query Intents: Automatic Generation

Many methods have been proposed

- Common ideas
  – Cluster top retrieved documents
    » By text, URL, keywords, etc.
  – Topic modeling on top retrieved documents
  – Pick diversified terms from search result page using existing document summarization tools
- None of them are convincing
  – E.g., not nearly as effective as Google suggestions
- Still an open research topic

14

14

Topic Outline

- Introduction
- Diversity Evaluation Metrics
- Implicit Methods
  - Maximum Marginal Relevance (MMR)
  - Learning to Rank for diversification
- **Explicit Methods**
  - Query intents (subtopics) discovery
  - xQuAD
  - PM-2

15

15

xQuAD

eXplicit Query Aspect Diversification (xQuAD)

- **Inputs**
  - An initial ranking
  - A set of query intents (e.g., from search engine suggestions)
- **Observation:** A single document may cover multiple intents
  - More likely to be true for faceted queries
- **Key idea:** Select documents that satisfy as many uncovered intents as possible
  - Provide maximum coverage and minimum redundancy
  - MMR did this implicitly – xQuAD does it explicitly

(Santos, et al., 2010)

16

16

xQuAD

$\textbf{xQuAD}(q, R, \tau, \lambda)$

$S \leftarrow \emptyset$
while $|S| < \tau$ **do**
  $d^* \leftarrow \arg\max_{d \in R \setminus S} \; (1-\lambda)\, \mathrm{P}(d|q) + \lambda\, \mathrm{P}(d, \bar{S}|q)$
  $R \leftarrow R \setminus \{d^*\}$
  $S \leftarrow S \cup \{d^*\}$
end while
return S

Relevance **Diversity**

Remove d^* from the initial ranking
Add d^* to the diversified ranking

R: Initial ranking (produced by some other method)

S: Diversified ranking (initially empty)

τ: Desired length of diversified ranking

λ: Balance between relevance and diversity

(Santos, et al., 2010)

Diversity Notation Alert

Earlier lectures defined q_i as the i^{th} term of query q

- q: search engines course
- q_2: 'engines'

This lecture defines q_i as the i^{th} intent of query q

- q: michael jordan
- q_2: The Berkeley professor Michael Jordan

This is how notation is used in most of the published work

- I'm sorry – don't be confused

xQuAD

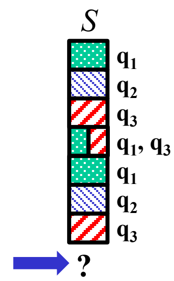
xQuAD's selection criteria: $\underset{\textbf{Relevance}}{(1-\lambda)\,\mathrm{P}(d|q)} + \underset{\textbf{Diversity}}{\lambda\,\mathrm{P}(d,\bar{S}|q)}$

Make some assumptions

- A query q has intents specified by subqueries $\{q_1, \ldots, q_k\}$
- Independence assumptions

xQuAD's diversity component becomes:

$$\mathrm{P}(d,\bar{S}|q) = \sum_{q_i \in Q} \left[\underset{\textbf{Intent weight}}{\mathrm{P}(q_i|q)}\ \underset{\textbf{Score of d for intent } q_i}{\mathrm{P}(d|q_i)}\ \underbrace{\prod_{d_j \in S} (1-\mathrm{P}(d_j|q_i))}_{\textbf{How well S already covers intent } q_i} \right]$$



19

19

xQuAD Example:

$$(1-\lambda)\,\mathrm{P}(d|q) + \lambda \sum_{q_i \in Q} \left[\mathrm{P}(q_i|q)\,\mathrm{P}(d|q_i) \prod_{d_j \in S} (1-\mathrm{P}(d_j|q_i)) \right]$$

$\mathbf{rank_{init}}$

d_1, 0.70

d_2, 0.69

d_3, 0.68

d_4, 0.67

d_5, 0.66

20

20

xQuAD Example:

$$(1-\lambda)\,\mathrm{P}(d|q)+\lambda\sum_{q_i\in Q}\Big[\mathrm{P}(q_i|q)\,\mathrm{P}(d|q_i)\prod_{d_j\in S}(1-\mathrm{P}(d_j|q_i))\Big]$$

p(q$_i$|q)=0.5

rank$_{init}$	**q$_1$**	**q$_2$**
d$_1$, 0.70,	0.7,	0.2
d$_2$, 0.69,	0.8,	0.1
d$_3$, 0.68,	0.6,	0.3
d$_4$, 0.67,	0.2,	0.7
d$_5$, 0.66,	0.3,	0.8

xQuAD Example:

$$(1-\lambda)\,\mathrm{P}(d|q)+\lambda\sum_{q_i\in Q}\Big[\mathrm{P}(q_i|q)\,\mathrm{P}(d|q_i)\prod_{d_j\in S}(1-\mathrm{P}(d_j|q_i))\Big]$$

p(q$_i$|q)=0.5

rank$_{init}$	**q$_1$**	**q$_2$**
d$_1$, 0.70,	$\boxed{0.7}$,	0.2
d$_2$, 0.69,	$\boxed{0.8}$,	0.1
d$_3$, 0.68,	$\boxed{0.6}$,	0.3
d$_4$, 0.67,	0.2,	$\boxed{0.7}$
d$_5$, 0.66,	0.3,	$\boxed{0.8}$

xQuAD Example:

$$(1-\lambda)\,\mathrm{P}(d|q) + \lambda \sum_{q_i \in Q} \left[\mathrm{P}(q_i|q)\,\mathrm{P}(d|q_i) \prod_{d_j \in S} (1 - \mathrm{P}(d_j|q_i)) \right]$$

$p(q_i|q)=0.5$
$\lambda=0.4$

$rank_{init}$	**q_1**	**q_2**	**xQuAD Scores**
d_1, 0.70,	0.7,	0.2	0.6×0.70 + 0.4×[½×0.7 + ½×0.2] = 0.600
d_2, 0.69,	0.8,	0.1	0.6×0.69 + 0.4×[½×0.8 + ½×0.1] = 0.594
d_3, 0.68,	0.6,	0.3	0.6×0.68 + 0.4×[½×0.6 + ½×0.3] = 0.588
d_4, 0.67,	0.2,	0.7	0.6×0.67 + 0.4×[½×0.2 + ½×0.7] = 0.582
d_5, 0.66,	0.3,	0.8	0.6×0.66 + 0.4×[½×0.3 + ½×0.8] = 0.616

23

23

xQuAD Example:

$$(1-\lambda)\,\mathrm{P}(d|q) + \lambda \sum_{q_i \in Q} \left[\mathrm{P}(q_i|q)\,\mathrm{P}(d|q_i) \prod_{d_j \in S} (1 - \mathrm{P}(d_j|q_i)) \right]$$

$p(q_i|q)=0.5$
$\lambda=0.4$

$rank_{init}$	**q_1**	**q_2**	**$rank_{div}$**
d_1, 0.70,	0.7,	0.2	d_5
d_2, 0.69,	0.8,	0.1	
d_3, 0.68,	0.6,	0.3	
d_4, 0.67,	0.2,	0.7	
d_5, 0.66,	0.3,	0.8	

24

24

xQuAD Example:

$$(1-\lambda)\,\mathrm{P}(d|q) + \lambda \sum_{q_i \in Q} \Big[\mathrm{P}(q_i|q)\,\mathrm{P}(d|q_i) \prod_{d_j \in S} (1 - \mathrm{P}(d_j|q_i))\Big]$$

p(q_i|q)=0.5
λ=0.4

$\text{rank}_{\text{init}}$	q_1	q_2	rank_{div}	xQuAD Scores
d_1, 0.70,	0.7,	0.2	d_5	0.6×0.70 + 0.4×[½×0.7(1-0.3) + ½×0.2(1-0.8)] = 0.526
d_2, 0.69,	0.8,	0.1		0.6×0.69 + 0.4×[½×0.8(1-0.3) + ½×0.1(1-0.8)] = 0.530
d_3, 0.68,	0.6,	0.3		0.6×0.68 + 0.4×[½×0.6(1-0.3) + ½×0.3(1-0.8)] = 0.504
d_4, 0.67,	0.2,	0.7		0.6×0.67 + 0.4×[½×0.2(1-0.3) + ½×0.7(1-0.8)] = 0.458
d_5, 0.66,	0.3,	0.8		

25

25

xQuAD Example:

$$(1-\lambda)\,\mathrm{P}(d|q) + \lambda \sum_{q_i \in Q} \Big[\mathrm{P}(q_i|q)\,\mathrm{P}(d|q_i) \prod_{d_j \in S} (1 - \mathrm{P}(d_j|q_i))\Big]$$

p(q_i|q)=0.5
λ=0.4

$\text{rank}_{\text{init}}$	q_1	q_2	rank_{div}
d_1, 0.70,	0.7,	0.2	d_5
d_2, 0.69,	0.8,	0.1	d_2
d_3, 0.68,	0.6,	0.3	
d_4, 0.67,	0.2,	0.7	
d_5, 0.66,	0.3,	0.8	

26

26

xQuAD Example:

$$(1-\lambda)\,\mathrm{P}(d|q) + \lambda \sum_{q_i \in Q} \left[\mathrm{P}(q_i|q)\,\mathrm{P}(d|q_i) \prod_{d_j \in S} (1 - \mathrm{P}(d_j|q_i)) \right]$$

$p(q_i|q)=0.5$
$\lambda=0.4$

$rank_{init}$	q_1	q_2	$rank_{div}$	xQuAD Scores
d_1, 0.70,	0.7,	0.2	d_5	0.6×0.70 + 0.4×[½×0.7(1-0.3)(1-0.8) + ½×0.2(1-0.8)(1-0.1)] = 0.4468
d_2, 0.69,	0.8,	0.1	d_2	
d_3, 0.68,	0.6,	0.3		0.6×0.68 + 0.4×[½×0.6(1-0.3)(1-0.8) + ½×0.3(1-0.8)(1-0.1)] = 0.4356
d_4, 0.67,	0.2,	0.7		0.6×0.67 + 0.4×[½×0.2(1-0.3)(1-0.8) + ½×0.7(1-0.8)(1-0.1)] = 0.4328
d_5, 0.66,	0.3,	0.8		

27

27

xQuAD Example:

$$(1-\lambda)\,\mathrm{P}(d|q) + \lambda \sum_{q_i \in Q} \left[\mathrm{P}(q_i|q)\,\mathrm{P}(d|q_i) \prod_{d_j \in S} (1 - \mathrm{P}(d_j|q_i)) \right]$$

$p(q_i|q)=0.5$
$\lambda=0.4$

$rank_{init}$	q_1	q_2	$rank_{div}$
d_1, 0.70,	0.7,	0.2	d_5
d_2, 0.69,	0.8,	0.1	d_2
d_3, 0.68,	0.6,	0.3	d_1
d_4, 0.67,	0.2,	0.7	
d_5, 0.66,	0.3,	0.8	

28

28

xQuAD Example:

$$(1-\lambda)\,\mathrm{P}(d|q) + \lambda \sum_{q_i \in Q} \left[\mathrm{P}(q_i|q)\,\mathrm{P}(d|q_i) \prod_{d_j \in S} (1 - \mathrm{P}(d_j|q_i)) \right]$$

$p(q_i|q)=0.5$
$\lambda=0.4$

$rank_{init}$	**q_1**	**q_2**	**$rank_{div}$**	**xQuAD Scores**
d_1, 0.70,	0.7,	0.2	d_5	
d_2, 0.69,	0.8,	0.1	d_2	
d_3, 0.68,	0.6,	0.3	d_1	0.6×0.68 + 0.4×[½×0.6(1-0.3)(1-0.8)(1-0.7) + ½×0.3(1-0.8)(1-0.1)(1-0.2)] = 0.4217
d_4, 0.67,	0.2,	0.7		0.6×0.67 + 0.4×[½×0.2(1-0.3)(1-0.8)(1-0.7) + ½×0.7(1-0.8)(1-0.1)(1-0.2)] = 0.4238
d_5, 0.66,	0.3,	0.8		

29

29

xQuAD Example:

$$(1-\lambda)\,\mathrm{P}(d|q) + \lambda \sum_{q_i \in Q} \left[\mathrm{P}(q_i|q)\,\mathrm{P}(d|q_i) \prod_{d_j \in S} (1 - \mathrm{P}(d_j|q_i)) \right]$$

$p(q_i|q)=0.5$
$\lambda=0.4$

$rank_{init}$	**q_1**	**q_2**	**$rank_{div}$**
d_1, 0.70,	0.7,	0.2	d_5
d_2, 0.69,	0.8,	0.1	d_2
d_3, 0.68,	0.6,	0.3	d_1
d_4, 0.67,	0.2,	0.7	d_4
d_5, 0.66,	0.3,	0.8	

30

30

xQuAD Example:

$$(1-\lambda)\,\mathrm{P}(d|q) + \lambda \sum_{q_i \in Q} \left[\mathrm{P}(q_i|q)\,\mathrm{P}(d|q_i) \prod_{d_j \in S} (1 - \mathrm{P}(d_j|q_i)) \right]$$

$rank_{init}$	q_1	q_2	$rank_{div}$
d_1, 0.70,	$\boxed{0.7,}$	0.2	d_5
d_2, 0.69,	$\boxed{0.8,}$	0.1	d_2
d_3, 0.68,	$\boxed{0.6,}$	0.3	d_1
d_4, 0.67,	0.2,	$\boxed{0.7}$	d_4
d_5, 0.66,	0.3,	$\boxed{0.8}$	

$p(q_i|q)=0.5$
$\lambda=0.4$

Repeat until all documents in the initial ranking are added to the diversified ranking

xQuAD Behavior

xQuAD prefers documents that satisfy multiple intents

$rank_{init}$

d_1, 0.70
d_2, 0.69
d_3, 0.68
d_4, 0.67
d_5, 0.66
d_6, 0.65
d_7, 0.64
d_8, 0.63

xQuAD Behavior

$p(q_i|q)=0.5$
$\lambda=0.4$

xQuAD prefers documents that satisfy multiple intents

$rank_{init}$	q_1	q_2
d_1, 0.70,	0.7,	0.1
d_2, 0.69,	0.7,	0.2
d_3, 0.68,	0.4,	0.5
d_4, 0.67,	0.5,	0.4
d_5, 0.66,	0.5,	0.5
d_6, 0.65,	0.5,	0.5
d_7, 0.64,	0.1,	0.7
d_8, 0.63,	0.1,	0.9

33

33

xQuAD Behavior

$p(q_i|q)=0.5$
$\lambda=0.4$

xQuAD prefers documents that satisfy multiple intents

$rank_{init}$	q_1	q_2
d_1, 0.70,	0.7,	0.1
d_2, 0.69,	0.7,	0.2
d_3, 0.68,	0.4,	0.5
d_4, 0.67,	0.5,	0.4
d_5, 0.66,	0.5,	0.5
d_6, 0.65,	0.5,	0.5
d_7, 0.64,	0.1,	0.7
d_8, 0.63,	0.1,	0.9

34

34

xQuAD Behavior

$p(q_i|q)=0.5$
$\lambda=0.4$

xQuAD prefers documents that satisfy multiple intents

$rank_{init}$	q_1	q_2	$rank_1$
d_1, 0.70,	0.7,	0.1	0.580
d_2, 0.69,	0.7,	0.2	0.594
d_3, 0.68,	0.4,	0.5	0.588
d_4, 0.67,	0.5,	0.4	0.582
d_5, 0.66,	0.5,	0.5	0.596
d_6, 0.65,	0.5,	0.5	0.590
d_7, 0.64,	0.1,	0.7	0.544
d_8, 0.63,	0.1,	0.9	0.578

Selecting the first document

35

35

xQuAD Behavior

$p(q_i|q)=0.5$
$\lambda=0.4$

xQuAD prefers documents that satisfy multiple intents

$rank_{init}$	q_1	q_2	$rank_1$	$rank_2$
d_1, 0.70,	0.7,	0.1	0.580	0.500
d_2, 0.69,	0.7,	0.2	0.594	0.504
d_3, 0.68,	0.4,	0.5	0.588	0.498
d_4, 0.67,	0.5,	0.4	0.582	0.492
d_5, 0.66,	0.5,	0.5	0.596	
d_6, 0.65,	0.5,	0.5	0.590	0.490
d_7, 0.64,	0.1,	0.7	0.544	0.464
d_8, 0.63,	0.1,	0.9	0.578	0.478

Selecting the second document

36

36

xQuAD Behavior

$p(q_i|q)=0.5$
$\lambda=0.4$

xQuAD prefers documents that satisfy multiple intents

	$rank_{init}$	q_1	q_2	$rank_1$	$rank_2$	$rank_3$
d_1	0.70	0.7	0.1	0.580	0.500	0.449
d_2	0.69	0.7	0.2	0.594	0.504	
d_3	0.68	0.4	0.5	0.588	0.498	0.460
d_4	0.67	0.5	0.4	0.582	0.492	0.449
d_5	0.66	0.5	0.5	0.596		
d_6	0.65	0.5	0.5	0.590	0.490	0.445
d_7	0.64	0.1	0.7	0.544	0.464	0.443
d_8	0.63	0.1	0.9	0.578	0.478	0.453

Selecting the third document

37

37

xQuAD Behavior

$p(q_i|q)=0.5$
$\lambda=0.4$

xQuAD prefers documents that satisfy multiple intents

	$rank_{init}$	q_1	q_2	$rank_1$	$rank_2$	$rank_3$	$rank_4$
d_1	0.70	0.7	0.1	0.580	0.500	0.449	0.437
d_2	0.69	0.7	0.2	0.594	0.504		
d_3	0.68	0.4	0.5	0.588	0.498	0.460	
d_4	0.67	0.5	0.4	0.582	0.492	0.449	0.427
d_5	0.66	0.5	0.5	0.596			
d_6	0.65	0.5	0.5	0.590	0.490	0.445	0.419
d_7	0.64	0.1	0.7	0.544	0.464	0.443	0.414
d_8	0.63	0.1	0.9	0.578	0.478	0.453	0.416

Selecting the fourth document

38

38

xQuAD Behavior

$p(q_i|q)=0.5$
$\lambda=0.4$

xQuAD prefers documents that satisfy multiple intents

	$rank_{init}$	q_1	q_2	$rank_1$	$rank_2$	$rank_3$	$rank_4$	$rank_5$
d_1,	0.70,	0.7,	0.1	0.580	0.500	0.449	0.437	
d_2,	0.69,	0.7,	0.2	0.594	0.504			
d_3,	0.68,	0.4,	0.5	0.588	0.498	0.460		
d_4,	0.67,	0.5,	0.4	0.582	0.492	0.449	0.427	0.419
d_5,	0.66,	0.5,	0.5	0.596				
d_6,	0.65,	0.5,	0.5	0.590	0.490	0.445	0.419	0.411
d_7,	0.64,	0.1,	0.7	0.544	0.464	0.443	0.414	0.410
d_8,	0.63,	0.1,	0.9	0.578	0.478	0.453	0.416	0.411

Selecting the fifth document

39

39

xQuAD Behavior

$p(q_i|q)=0.5$
$\lambda=0.4$

xQuAD prefers documents that satisfy multiple intents

	$rank_{init}$	q_1	q_2	$rank_1$	$rank_2$	$rank_3$	$rank_4$	$rank_5$
d_1,	0.70,	0.7,	0.1	0.580	0.500	0.449	0.437	
d_2,	0.69,	0.7,	0.2	0.594	0.504			
d_3,	0.68,	0.4,	0.5	0.588	0.498	0.460		
d_4,	0.67,	0.5,	0.4	0.582	0.492	0.449	0.427	0.419
d_5,	0.66,	0.5,	0.5	0.596				
d_6,	0.65,	0.5,	0.5	0.590	0.490	0.445	0.419	0.411
d_7,	0.64,	0.1,	0.7	0.544	0.464	0.443	0.414	0.410
d_8,	0.63,	0.1,	0.9	0.578	0.478	0.453	0.416	0.411

q_2 is covered by mediocre documents that also cover q_1, so d_7 and d_8 don't get promoted

40

40

xQuAD

Santos, et al investigated three methods of weighting queries

- Uniform weights: 1 / |Q|
- A method similar to the CRCS resource ranking algorithm
  - In a later lecture we will cover ReDDE, a similar algorithm
  - Related to the number of top-ranked documents q_i matches
- Based on the relative number of documents it matches in a commercial search engine
  - $n_w(q_i)$ on an earlier slide

Uniform was the most effective on TREC 2009 data

- Perhaps because TREC intents are all weighted equally

(Santos, et al., 2010)

41

41

xQuAD

Effectiveness when using TREC intents (the gold standard)

	α-NDCG			IA-P			
	@5	@10	@100	@5	@10	@100	
BM25	0.159	0.186	0.288	0.075	0.071	**0.059**	Initial ranking
+MMR	0.120	0.150	0.224	0.056	0.058	0.039	
+Q-Filter	0.159	0.186	0.286	0.075	0.071	0.057	
+IA-Select	0.110	0.119	0.180	0.043	0.037	0.023	
+xQuAD$_u$	**0.208**	**0.227**	**0.324**	**0.080**	**0.075**	0.056	
DPH	0.198	0.212	0.304	**0.109**	**0.106**	**0.062**	Initial ranking
+MMR	0.195	0.211	0.303	0.105	0.103	**0.062**	
+Q-Filter	0.198	0.212	0.303	**0.109**	**0.106**	0.060	
+IA-Select	0.148	0.157	0.203	0.077	0.071	0.023	
+xQuAD$_u$	**0.208**	**0.243**	**0.334**	0.097	0.096	0.061	

ClueWeb09-B (50 queries from 2009)

(Santos, et al., 2010)

42

42

xQuAD

Effectiveness when using search engine queries as intents

		related sub-queries					
		α-NDCG			IA-P		
	WSE	@5	@10	@100	@5	@10	@100
BM25		0.159	**0.186**	**0.288**	0.075	0.071	**0.059**
+xQuAD$_u$	A	0.154	0.184	0.282	0.070	0.072	0.057
+xQuAD$_u$	B	0.154	0.182	0.279	0.073	**0.076**	0.054
+xQuAD$_u$	C	**0.161**	0.182	0.285	**0.076**	**0.076**	0.057
DPH		0.198	**0.212**	0.304	**0.109**	**0.106**	**0.062**
+xQuAD$_u$	A	0.164	0.189	0.288	0.086	0.083	0.056
+xQuAD$_u$	B	0.186	0.205	0.295	0.090	0.082	0.057
+xQuAD$_u$	C	**0.206**	0.209	**0.307**	0.108	0.090	**0.062**

Initial ranking

Initial ranking

ClueWeb09-B (50 queries from 2009)

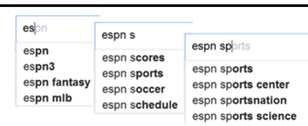
Related queries (from web search engines A, B, and C) provide little benefit

(Santos, et al., 2010)

43

43

xQuAD



Effectiveness when using search engine queries as intents

		suggested sub-queries					
		α-NDCG			IA-P		
	WSE	@5	@10	@100	@5	@10	@100
BM25		0.159	**0.186**	0.288	0.075	**0.071**	**0.059**
+xQuAD$_u$	A	**0.171**	**0.186**	**0.291**	0.082	**0.071**	0.053
+xQuAD$_u$	B	0.129	0.158	0.261	0.065	0.067	0.052
+xQuAD$_u$	C	0.163	0.184	0.287	**0.084**	0.069	0.053
DPH		0.198	0.212	0.304	**0.109**	**0.106**	**0.062**
+xQuAD$_u$	A	**0.215**	0.222	0.313	0.108	0.088	0.055
+xQuAD$_u$	B	0.162	0.189	0.281	0.088	0.085	0.055
+xQuAD$_u$	C	0.201	**0.236**	**0.320**	0.093	0.092	0.059

Initial ranking

Initial ranking

ClueWeb09-B (50 queries from 2009)

Suggested queries (from some engines) are more effective

(Santos, et al., 2010)

44

44

xQuAD

Effectiveness when using search engine queries (Bing) as intents

		$\mathcal{S}_q$	ERR-IA				α-nDCG				
			@20	–	=	+	@20	–	=	+	
	DPH		0.253				0.364				**Initial ranking**
	+MMR		$0.253^{\circ\circ}$	55	30	60	$0.367^{\circ\triangledown}$	56	28	61	
	+PC	BS	$0.256^{\blacktriangle\circ}$	25	58	62	$0.375^{\blacktriangle\triangledown}$	29	55	61	
	+IA-Select	DZ	$0.250^{\circ\circ}$	67	12	66	$0.356^{\circ\triangledown}$	70	12	63	
	+xQuAD	BS	$\underline{0.281}^{\circ}$	40	24	81	$\underline{0.402}^{\blacktriangle}$	37	24	84	
LeToR	LambdaMART		0.337				0.464				**Initial ranking**
	+MMR		$0.338^{\circ\circ}$	69	20	56	$0.466^{\circ\circ}$	69	20	56	
	+PC	BS	$0.339^{\blacktriangle\circ}$	27	52	66	$0.472^{\blacktriangle\circ}$	32	45	68	
	+IA-Select	DZ	$0.217^{\blacktriangledown\blacktriangledown}$	93	13	39	$0.329^{\blacktriangledown\blacktriangledown}$	98	13	34	
	+xQuAD	BS	$\underline{\mathbf{0.351}}^{\vartriangle}$	43	24	78	$\underline{\mathbf{0.479}}^{\blacktriangle}$	42	23	80	

(ClueWeb09-B, 145 TREC queries from 2009-2011)

(Santos, et al., 2010)

xQuAD

xQuAD characteristics

- Requires a set of explicit query intents
- Allows a single document to satisfy multiple query intents
  - Those documents are more likely to be ranked highly
- The diversified ranking covers each intent <u>equally</u>
  - Weighting is possible, but uniform weights were best
- Each query q requires running queries q and $\{q_1, \ldots, q_k\}$
  - 7-10 queries in Santos, et al's experiments
  - Thus a little expensive computationally
- One of the more effective methods available today

Topic Outline

- Introduction
- Diversity Evaluation Metrics
- Implicit Methods
  - Maximum Marginal Relevance (MMR)
  - Learning to Rank for diversification
- **Explicit Methods**
  - Query intents (subtopics) discovery
  - xQuAD
  - PM-2

47

47

PM-2

Proportionality Model 2 (PM-2)

- **Inputs**
  - An initial ranking
  - A set of query intents (e.g., from search engine suggestions)
- **Observation:** A high Recall algorithm for discovering query intents would find many rare or unpopular intents
  - They shouldn't get equal coverage in a diversified ranking
- **Key idea:** The number of documents for each subtopic should be proportional to each subtopic's popularity
  - Minimizes redundancy
  - Coverage of subtopics may better match a person's expectation

(Dang and Croft, 2012)

48

48

PM-2

PM-2 is an adaptation of the Sainte-Laguë method of assigning proportional representation in elections

- It is a greedy algorithm
- At each rank r
  - Select the query intent q_i that must be covered next to maintain proportional coverage of intents in the ranking
  - Select a document d that covers intent q_i
    - » And, maybe also covers other query intents
  - Remove d from the initial ranking R
  - Assign d to the diversified ranking S

A document may cover multiple intents to varying degrees

(Dang and Croft, 2012)

PM-2

$\forall i, v_i = \dfrac{\text{Diversified Ranking Size}}{\text{Number Of Intents}}, s_i = 0$ — **v_i: desired ranks in S for q_i**

s_i: ranks in S assigned to q_i now

for all $s \in S$ — **s: a rank in S to be filled**

$\forall q_i \in Q, qt[i] = \dfrac{v_i}{2s_i + 1}$ — **Priority of each intent now**

$i^* = \arg\max_i qt[i]$ — **i*: The intent to cover now**

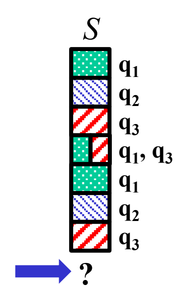
$$d^* = \arg\max_{d_j \in R} \Big(\underbrace{\lambda\, qt[i^*] p(d_j|q_{i^*})}_{\text{Covers } q_{i^*}} + (1-\lambda) \underbrace{\textstyle\sum_{i \neq i^*} qt[i] p(d_j|q_i)}_{\text{Also covers other intents}}\Big)$$

$S = S \cup \{d^*\}$

$R = R \setminus \{d^*\}$

$\forall q_i \in Q, s_i = s_i + \dfrac{p(d^*|q_i)}{\sum_{q_j \in Q} p(d^*|q_j)}$ — **Update coverage of each intent**

end for



(Dang and Croft, 2012)

PM-2 Example

$\forall i, v_i = \frac{\text{Diversified Ranking Size}}{\text{Number Of Intents}}, s_i = 0$

for all $s \in S$

$\quad \forall q_i \in Q, qt[i] = \frac{v_i}{2s_i + 1}$

$\quad i^* = \arg\max_i qt[i]$

$\quad d^* = \arg\max_{d_j \in R} \left(\lambda\, qt[i^*] p(d_j|q_{i^*}) + (1-\lambda) \sum_{i \neq i^*} qt[i] p(d_j|q_i)\right)$

$\quad S = S \cup \{d^*\}$

$\quad R = R \setminus \{d^*\}$

$\quad \forall q_i \in Q, s_i = s_i + \frac{p(d^*|q_i)}{\sum_{q_j \in Q} p(d^*|q_j)}$

end for

$rank_{init}$	q_1	q_2
d_1, 0.70,	0.7,	0.2
d_2, 0.69,	0.8,	0.1
d_3, 0.68,	0.6,	0.3
d_4, 0.67,	0.2,	0.7
d_5, 0.66,	0.3,	0.8

Initialization

λ=0.6

Example assumptions

- $p(q_i|q)$=0.5
- Ranking depth=8

v[1] = 0.5×8 = 4 “votes”
v[2] = 0.5×8 = 4

s[1] = 0 “ranks (slots) assigned so far”
s[2] = 0

PM-2 Example

$\forall i, v_i = \frac{\text{Diversified Ranking Size}}{\text{Number Of Intents}}, s_i = 0$

for all $s \in S$

$\quad \forall q_i \in Q, qt[i] = \frac{v_i}{2s_i + 1}$

$\quad i^* = \arg\max_i qt[i]$

$\quad d^* = \arg\max_{d_j \in R} \left(\lambda\, qt[i^*] p(d_j|q_{i^*}) + (1-\lambda) \sum_{i \neq i^*} qt[i] p(d_j|q_i)\right)$

$\quad S = S \cup \{d^*\}$

$\quad R = R \setminus \{d^*\}$

$\quad \forall q_i \in Q, s_i = s_i + \frac{p(d^*|q_i)}{\sum_{q_j \in Q} p(d^*|q_j)}$

end for

$rank_{init}$	q_1	q_2
d_1, 0.70,	0.7,	0.2
d_2, 0.69,	0.8,	0.1
d_3, 0.68,	0.6,	0.3
d_4, 0.67,	0.2,	0.7
d_5, 0.66,	0.3,	0.8

Quotient scores

λ=0.6

qt[1] = 4 / (2×0+1) = 4 “quotients (priorities)”
qt[2] = 4 / (2×0+1) = 4

i^* = 1 “selected intent”

PM-2 Example

$\forall i, v_i = \frac{\text{Diversified Ranking Size}}{\text{Number Of Intents}}, s_i = 0$
for all $s \in S$
  $\forall q_i \in Q, qt[i] = \frac{v_i}{2s_i + 1}$
  $i^* = \arg\max_i qt[i]$
  $d^* = \arg\max_{d_j \in R} \left(\lambda\, qt[i^*] p(d_j|q_{i^*}) + (1-\lambda) \sum_{i \neq i^*} qt[i] p(d_j|q_i)\right)$
  $S = S \cup \{d^*\}$
  $R = R \setminus \{d^*\}$
  $\forall q_i \in Q, s_i = s_i + \frac{p(d^*|q_i)}{\sum_{q_j \in Q} p(d^*|q_j)}$
end for

λ=0.6

$rank_{init}$	q_1	q_2	PM2 Scores
d_1, 0.70,	0.7,	0.2	0.6×4×0.7 + 0.4×4×0.2 = 2.00
d_2, 0.69,	0.8,	0.1	0.6×4×0.8 + 0.4×4×0.1 = 2.08
d_3, 0.68,	0.6,	0.3	0.6×4×0.6 + 0.4×4×0.3 = 1.92
d_4, 0.67,	0.2,	0.7	0.6×4×0.2 + 0.4×4×0.7 = 1.60
d_5, 0.66,	0.3,	0.8	0.6×4×0.3 + 0.4×4×0.8 = 2.00

PM-2 Example

$\forall i, v_i = \frac{\text{Diversified Ranking Size}}{\text{Number Of Intents}}, s_i = 0$
for all $s \in S$
  $\forall q_i \in Q, qt[i] = \frac{v_i}{2s_i + 1}$
  $i^* = \arg\max_i qt[i]$
  $d^* = \arg\max_{d_j \in R} \left(\lambda\, qt[i^*] p(d_j|q_{i^*}) + (1-\lambda) \sum_{i \neq i^*} qt[i] p(d_j|q_i)\right)$
  $S = S \cup \{d^*\}$
  $R = R \setminus \{d^*\}$
  $\forall q_i \in Q, s_i = s_i + \frac{p(d^*|q_i)}{\sum_{q_j \in Q} p(d^*|q_j)}$
end for

λ=0.6

$rank_{init}$	q_1	q_2	$rank_{div}$	Slots assigned
d_1, 0.70,	0.7,	0.2	d_2	s[1] += 0.8/(0.8+0.1) = 0.89 s[2] += 0.1/(0.8+0.1) = 0.11
d_2, 0.69,	0.8,	0.1		
d_3, 0.68,	0.6,	0.3		
d_4, 0.67,	0.2,	0.7		
d_5, 0.66,	0.3,	0.8		

PM-2 Example

$\forall i, v_i = \frac{\text{Diversified Ranking Size}}{\text{Number Of Intents}}, s_i = 0$
for all $s \in S$
$\forall q_i \in Q, qt[i] = \frac{v_i}{2s_i + 1}$
$i^* = \arg\max_i qt[i]$
$d^* = \arg\max_{d_j \in R} \left(\lambda\, qt[i^*] p(d_j|q_{i^*}) + (1-\lambda) \sum_{i \neq i^*} qt[i] p(d_j|q_i)\right)$
$S = S \cup \{d^*\}$
$R = R \setminus \{d^*\}$
$\forall q_i \in Q, s_i = s_i + \frac{p(d^*|q_i)}{\sum_{q_j \in Q} p(d^*|q_j)}$
end for

$rank_{init}$	q_1	q_2	$rank_{div}$
d_1, 0.70,	0.7,	0.2	d_2
d_2, 0.69,	0.8,	0.1	
d_3, 0.68,	0.6,	0.3	
d_4, 0.67,	0.2,	0.7	
d_5, 0.66,	0.3,	0.8	

Quotient scores

λ=0.6

qt[1] = 4 / (2×0.89+1) = 1.44 “quotients (priorities)”
qt[2] = 4 / (2×0.11+1) = 3.27

i^* = 2 “selected intent”

55

55

PM-2 Example

$\forall i, v_i = \frac{\text{Diversified Ranking Size}}{\text{Number Of Intents}}, s_i = 0$
for all $s \in S$
$\forall q_i \in Q, qt[i] = \frac{v_i}{2s_i + 1}$
$i^* = \arg\max_i qt[i]$
$d^* = \arg\max_{d_j \in R} \left(\lambda\, qt[i^*] p(d_j|q_{i^*}) + (1-\lambda) \sum_{i \neq i^*} qt[i] p(d_j|q_i)\right)$
$S = S \cup \{d^*\}$
$R = R \setminus \{d^*\}$
$\forall q_i \in Q, s_i = s_i + \frac{p(d^*|q_i)}{\sum_{q_j \in Q} p(d^*|q_j)}$
end for

$rank_{init}$	q_1	q_2	$rank_{div}$	PM2 Scores
d_1, 0.70,	0.7,	0.2	d_2	0.6×3.27×0.2 + 0.4×1.44×0.7 = 0.80
d_2, 0.69,	0.8,	0.1		
d_3, 0.68,	0.6,	0.3		0.6×3.27×0.3 + 0.4×1.44×0.6 = 0.93
d_4, 0.67,	0.2,	0.7		0.6×3.27×0.7 + 0.4×1.44×0.2 = 1.49
d_5, 0.66,	0.3,	0.8		0.6×3.27×0.8 + 0.4×1.44×0.3 = 1.74

λ=0.6

56

56

PM-2 Example

$rank_{init}$	q_1	q_2	$rank_{div}$
d_1, 0.70,	0.7,	0.2	d_2
d_2, 0.69,	0.8,	0.1	d_5
d_3, 0.68,	0.6,	0.3	
d_4, 0.67,	0.2,	0.7	
d_5, 0.66,	0.3,	0.8	

Slots assigned

λ=0.6

s[1] += 0.3/(0.3+0.8) = 1.16
s[2] += 0.8/(0.3+0.8) = 0.84

57

57

PM-2 Example

$rank_{init}$	q_1	q_2	$rank_{div}$
d_1, 0.70,	0.7,	0.2	d_2
d_2, 0.69,	0.8,	0.1	d_5
d_3, 0.68,	0.6,	0.3	
d_4, 0.67,	0.2,	0.7	
d_5, 0.66,	0.3,	0.8	

Quotient scores

λ=0.6

qt[1] = 4 / (2×1.16+1) = 1.20 "quotients (priorities)"
qt[2] = 4 / (2×0.84+1) = 1.49

$i^* = 2$ "selected intent"

58

58

PM-2 Example

$\forall i, v_i = \frac{\text{Diversified Ranking Size}}{\text{Number Of Intents}}, s_i = 0$
for all $s \in S$
$\forall q_i \in Q, qt[i] = \frac{v_i}{2s_i + 1}$
$i^* = \arg\max_i qt[i]$
$d^* = \arg\max_{d_j \in R} \left(\lambda\, qt[i^*] p(d_j|q_{i^*}) + (1-\lambda) \sum_{i \neq i^*} qt[i] p(d_j|q_i)\right)$
$S = S \cup \{d^*\}$
$R = R \setminus \{d^*\}$
$\forall q_i \in Q, s_i = s_i + \frac{p(d^*|q_i)}{\sum_{q_j \in Q} p(d^*|q_j)}$
end for

$\lambda=0.6$

$rank_{init}$	q_1	q_2	$rank_{div}$	PM2 Scores
d_1, 0.70,	0.7,	0.2	d_2	0.6×1.49×0.2 + 0.4×1.20×0.7 = 0.52
d_2, 0.69,	0.8,	0.1	d_5	
d_3, 0.68,	0.6,	0.3		0.6×1.49×0.3 + 0.4×1.20×0.6 = 0.56
d_4, 0.67,	0.2,	0.7		0.6×1.49×0.7 + 0.4×1.20×0.2 = 0.72
d_5, 0.66,	0.3,	0.8		

59

PM-2 Example

$\forall i, v_i = \frac{\text{Diversified Ranking Size}}{\text{Number Of Intents}}, s_i = 0$
for all $s \in S$
$\forall q_i \in Q, qt[i] = \frac{v_i}{2s_i + 1}$
$i^* = \arg\max_i qt[i]$
$d^* = \arg\max_{d_j \in R} \left(\lambda\, qt[i^*] p(d_j|q_{i^*}) + (1-\lambda) \sum_{i \neq i^*} qt[i] p(d_j|q_i)\right)$
$S = S \cup \{d^*\}$
$R = R \setminus \{d^*\}$
$\forall q_i \in Q, s_i = s_i + \frac{p(d^*|q_i)}{\sum_{q_j \in Q} p(d^*|q_j)}$
end for

$\lambda=0.6$

$rank_{init}$	q_1	q_2	$rank_{div}$	Slots assigned
d_1, 0.70,	0.7,	0.2	d_2	s[1] = 1.16+0.2/(0.2+0.7) = 1.38 s[2] = 0.84+0.7/(0.2+0.7) = 1.62
d_2, 0.69,	0.8,	0.1	d_5	
d_3, 0.68,	0.6,	0.3	d_4	
d_4, 0.67,	0.2,	0.7		
d_5, 0.66,	0.3,	0.8		

60

PM-2 Example

$\forall i, v_i = \frac{\text{Diversified Ranking Size}}{\text{Number Of Intents}}, s_i = 0$
for all $s \in S$
  $\forall q_i \in Q, qt[i] = \frac{v_i}{2s_i + 1}$
  $i^* = \arg\max_i qt[i]$
  $d^* = \arg\max_{d_j \in R} \left(\lambda\, qt[i^*] p(d_j|q_{i^*}) + (1-\lambda) \sum_{i \neq i^*} qt[i] p(d_j|q_i)\right)$
  $S = S \cup \{d^*\}$
  $R = R \setminus \{d^*\}$
  $\forall q_i \in Q, s_i = s_i + \frac{p(d^*|q_i)}{\sum_{q_j \in Q} p(d^*|q_j)}$
end for

λ=0.6

$rank_{init}$	q_1	q_2	$rank_{div}$
d_1, 0.70,	0.7,	0.2	d_2
d_2, 0.69,	0.8,	0.1	d_5
d_3, 0.68,	0.6,	0.3	d_4
d_4, 0.67,	0.2,	0.7	
d_5, 0.66,	0.3,	0.8	

Quotient scores

qt[1] = 4 / (2×1.38+1) = 1.06 "quotients (priorities)"
qt[2] = 4 / (2×1.62+1) = 0.95

$i^* = 1$ "selected intent"

61

61

PM-2 Example

$\forall i, v_i = \frac{\text{Diversified Ranking Size}}{\text{Number Of Intents}}, s_i = 0$
for all $s \in S$
  $\forall q_i \in Q, qt[i] = \frac{v_i}{2s_i + 1}$
  $i^* = \arg\max_i qt[i]$
  $d^* = \arg\max_{d_j \in R} \left(\lambda\, qt[i^*] p(d_j|q_{i^*}) + (1-\lambda) \sum_{i \neq i^*} qt[i] p(d_j|q_i)\right)$
  $S = S \cup \{d^*\}$
  $R = R \setminus \{d^*\}$
  $\forall q_i \in Q, s_i = s_i + \frac{p(d^*|q_i)}{\sum_{q_j \in Q} p(d^*|q_j)}$
end for

λ=0.6

$rank_{init}$	q_1	q_2	$rank_{div}$	PM2 Scores
d_1, 0.70,	0.7,	0.2	d_2	0.6×1.06×0.7 + 0.4×0.95×0.2 = 0.52
d_2, 0.69,	0.8,	0.1	d_5	
d_3, 0.68,	0.6,	0.3	d_4	0.6×1.06×0.6 + 0.4×0.95×0.3 = 0.50
d_4, 0.67,	0.2,	0.7		
d_5, 0.66,	0.3,	0.8		

62

62

PM-2 Example

$\forall i, v_i = \frac{\text{Diversified Ranking Size}}{\text{Number Of Intents}}, s_i = 0$
for all $s \in S$
  $\forall q_i \in Q, qt[i] = \frac{v_i}{2s_i + 1}$
  $i^* = \arg\max_i qt[i]$
  $d^* = \arg\max_{d_j \in R} \left(\lambda\, qt[i^*] p(d_j|q_{i^*}) + (1-\lambda) \sum_{i \neq i^*} qt[i] p(d_j|q_i)\right)$
  $S = S \cup \{d^*\}$
  $R = R \setminus \{d^*\}$
  $\forall q_i \in Q, s_i = s_i + \frac{p(d^*|q_i)}{\sum_{q_j \in Q} p(d^*|q_j)}$
end for

$\lambda=0.6$

$rank_{init}$	q_1	q_2	$rank_{div}$
d_1, 0.70,	0.7,	0.2	d_2
d_2, 0.69,	0.8,	0.1	d_5
d_3, 0.68,	0.6,	0.3	d_4
d_4, 0.67,	0.2,	0.7	d_1
d_5, 0.66,	0.3,	0.8	

Slots assigned

s[1] = 1.38+0.7/(0.7+0.2) = 2.16
s[2] = 1.62+0.2/(0.7+0.2) = 1.84

63

63

PM-2 Example

$\forall i, v_i = \frac{\text{Diversified Ranking Size}}{\text{Number Of Intents}}, s_i = 0$
for all $s \in S$
  $\forall q_i \in Q, qt[i] = \frac{v_i}{2s_i + 1}$
  $i^* = \arg\max_i qt[i]$
  $d^* = \arg\max_{d_j \in R} \left(\lambda\, qt[i^*] p(d_j|q_{i^*}) + (1-\lambda) \sum_{i \neq i^*} qt[i] p(d_j|q_i)\right)$
  $S = S \cup \{d^*\}$
  $R = R \setminus \{d^*\}$
  $\forall q_i \in Q, s_i = s_i + \frac{p(d^*|q_i)}{\sum_{q_j \in Q} p(d^*|q_j)}$
end for

$\lambda=0.6$

$rank_{init}$	q_1	q_2	$rank_{div}$
d_1, 0.70,	0.7,	0.2	d_2
d_2, 0.69,	0.8,	0.1	d_5
d_3, 0.68,	0.6,	0.3	d_4
d_4, 0.67,	0.2,	0.7	d_1
d_5, 0.66,	0.3,	0.8	

Repeat until all documents in the initial ranking are added to the diversified ranking

64

64

xQuAD vs PM-2

xQuAD picks a document that covers multiple intents

- Give higher weight to intents that need coverage

PM2 picks the intent that most needs to be covered

- Then it picks a document that covers that intent
- There is extra credit for covering other intents, too

$rank_{init}$	q_1	q_2	$rank_P$	$rank_x$
d_1, 0.70,	0.7,	0.2	d_2	d_5
d_2, 0.69,	0.8,	0.1	d_5	d_2
d_3, 0.68,	0.6,	0.3	d_4	d_1
d_4, 0.67,	0.2,	0.7	d_1	d_4
d_5, 0.66,	0.3,	0.8	d_3	d_3

Notice the choices that each algorithms makes **λ_P=0.6, λ_x=0.4**

65

xQuAD vs PM-2

xQuAD picks a document that covers multiple intents

- Give higher weight to intents that need coverage

PM2 picks the intent that most needs to be covered

- Then it picks a document that covers that intent
- There is extra credit for covering other intents, too

$rank_{init}$	q_1	q_2	$rank_P$	$rank_x$
d_1, 0.70,	0.7,	0.1	d_2	d_5
d_2, 0.69,	0.7,	0.2	d_8	d_2
d_3, 0.68,	0.4,	0.5	d_5	d_3
d_4, 0.67,	0.5,	0.4	d_6	d_1
d_5, 0.66,	0.5,	0.5	d_1	d_4
d_6, 0.65,	0.5,	0.5	d_7	d_6
d_7, 0.64,	0.1,	0.7	d_4	d_7
d_8, 0.63,	0.1,	0.9	d_3	d_8

Notice the choices that each algorithms makes **λ_P=0.6, λ_x=0.4**

66

PM-2

Dang and Croft use uniform weighting of query intents

- Santos, et al found it to be most effective

(Dang and Croft, 2012)

PM-2

How well does it work?

		α-NDCG	Win/Loss	ERR-IA	Prec-IA
			WT-2009		
Sub-topics	Query-likelihood	0.2979		0.1953	0.1146
	MMR	0.2963	16/19	0.1922	**0.1221**
	xQuAD	$0.3300_{Q,M}$	23/15	$0.2207_{Q,M}$	0.1190
	PM-1	0.3076	18/17	0.2027	0.1140
	PM-2	$\mathbf{0.3473}^{P}$	19/19	$\mathbf{0.2407}^{P}$	0.1197
Suggestions	Query-likelihood	0.2875		0.1895	0.1095
	MMR	0.2926	16/15	0.1919	0.1108
	xQuAD	0.2995	14/19	0.1973	0.1089
	PM-1	0.2870	16/18	0.1830	0.0929^{X}
	PM-2	**0.3200**	17/19	**0.2139**	$\mathbf{0.1123}^{P}$
WT-2009 Best (uogTrDYCcsB) [10]		0.3081	N/A	0.1922	N/A

ClueWeb09-B (50 queries from 2009)

Evaluated at 50 documents (e.g., α-NDCG@50)

(Dang and Croft, 2012)

PM-2

How well does it work?

		α-NDCG	Win/Loss	ERR-IA	Prec-IA
			WT-2010		
Sub-topics	Query-likelihood	0.3236		0.2081	0.1713
	MMR	0.3349_{Q}	19/14	0.2161	0.1740
	xQuAD	$0.4074_{Q,M}$	29/14	$0.2671_{Q,M}$	0.2028
	PM-1	$0.4323^{X}_{Q,M}$	32/13	$0.3071^{X}_{Q,M}$	0.1827
	PM-2	$\mathbf{0.4546}^{X,P}_{Q,M}$	34/10	$\mathbf{0.3271}^{X}_{Q,M}$	**0.2030**
Suggestions	Query-likelihood	0.3268		0.2131	0.1730
	MMR	0.3361_{Q}	17/14	0.2206	0.1746
	xQuAD	$0.3582_{Q,M}$	31/6	$0.2372_{Q,M}$	0.1785
	PM-1	0.3664^{X}	25/15	0.2409	0.1654
	PM-2	$\mathbf{0.4374}^{X,P}_{Q,M}$	33/10	$\mathbf{0.3087}^{X,P}_{Q,M}$	**0.1841**
WT-2010 Best (uogTrB67xS) [11]		0.4178	N/A	0.2980	N/A

ClueWeb09-B (50 queries from 2010)

Evaluated at 50 documents (e.g., α-NDCG@50)

(Dang and Croft, 2012)

PM-2

PM-2 characteristics

- Requires a set of explicit query intents
- Allows a single document to satisfy multiple query intents
  - – Those documents are more likely to be ranked highly
- The diversified ranking covers each intent proportionally
  - – Weighting is possible, but uniform weights were best
- Each query q requires running queries q and $\{q_1, \ldots, q_k\}$
  - – Presumably 7-10 queries in Dang & Croft's experiments
  - – Thus a little expensive computationally
- Relative performance vs. xQuAD depends on detailed datasets and explicit topics.

Implicit vs. Explicit Diversification Algorithms

Explicit methods are more effective than MMR, but they need to know query intents

- Obtaining intents from web search engines is unsatisfying
  - Reliance on another organization
  - Perhaps the query suggestions don't work well for your task

R-LeToR works the best with supervision

- xQuAD and PM-2 are both unsupervised (but they have extra information – intents)

Ideally, we want the best characteristics of each method

- No need for external resources
- Make use of subtopics when available
- Fully utilize supervision

Summary

Methods of identifying query intents

- Commercial search engine suggestions
  - Query suggestions > related queries
- Still an open problem
  - Very hard without search log

Summary

Diversification algorithms

- Maximum Marginal Relevance (MMR)
- Relational Learning to Rank (R-LeToR)
- xQuAD
- PM-2

Characteristics

- MMR: Implicit, unsupervised, penalizes redundancy
- R-LeToR: Implicit, supervised, features to model redundancy
- xQuAD: Explicit, unsupervised, penalizes redundancy
- PM-2: Explicit, unsupervised, enforces proportionality

Summary

Performance: R-LeToR > PM-2 ≈ xQuAD > MMR

Topic Outline

- **Introduction**
- **Diversity Evaluation Metrics**
- **Implicit Methods**
  – Maximum Marginal Relevance (MMR)
  – Learning to Rank for diversification
- **Explicit Methods**
  – Query intents (subtopics) discovery
  – xQuAD
  – PM-2

For More Information

- R. Agrawal, S. Gollapudi, A. Halverson, and S. Ieong. Diversifying search results. In WSDM '09 Proceedings. 2009.
- J. Carbonell and J. Goldstein. The use of MMR, diversity-based reranking for reordering documents and producing summaries. Proceedings for SIGIR '98. 1998.
- C. L.A. Clarke, M. Kolla, G. V. Cormack, O. Vechtomova, A. Ashkann, S. Büttcher, and I. MacKinnon. Novelty and diversity in information retrieval evaluation. In SIGIR '08 Proceedings. 2008.
- V. Dang and W.B. Croft. Diversity by proportionality: An election-based approach to search result diversification. In Proceedings of SIGIR '12. 2012.
- V. Dang and W. B. Croft. Term-level search result diversification. Proceedings of SIGIR '13. 2013.
- D. Lawrie, W.B. Croft, and A. Rosenberg. Finding topic words for hierarchical summarization. In Proceedings of SIGIR. 2001.
- R. L. T. Santos, C. Macdonald, and I. Ounis. Exploiting query reformulations for web search result diversification. Proceedings of WWW 2010. 2010.
- R. L. T. Santos. Explicit web search result diversification. Ph.D. dissertation. 2013.
- Y. Zhu , Y. Lan, J. Guo, X. Cheng and S. Niu. Learning for search result diversification. SIGIR 2014. 2014.