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

Diversity

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

1

Introduction

The methods covered until now assume that each document is <u>independent</u> of other documents in the ranking

- This is a simple approach
- Often it is effective … but not always

When does it produce poor results?

- A query may have multiple interpretations or <u>user intents</u>
- If the ranking covers <u>just one intent</u>
  … and it is <u>the wrong intent</u>
  … the user will not find any relevant documents
  … extremely bad case, must avoid!

2

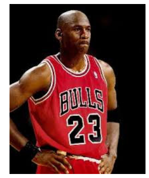
2

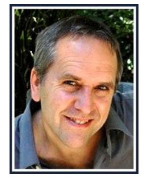
Introduction

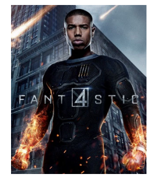
Query: Michael Jordan

Which interpretation is intended?

- The basketball player?
- The Berkeley professor?
- The former CEO of Pepsi, CBS & EDS?
- The actor?
- The State Farm insurance agent?







The first four are well-known

- The last is important to his customers

Introduction

<u>Most</u> short queries have multiple <u>interpretations</u>

- Jamie Callan
- java
- jaguar
- avp
- …





Introduction

Longer queries also have multiple interpretations

- **Query:** arizona game and fish
  - Arizona Game and Fish Department home page?
  - Regulations for hunting and fishing in Arizona?
  - The Arizona Fishing report site?
  - Guides and outfitters for hunting trips in Arizona?
- **Query:** discovery channel store
  - The Discovery Channel store homepage?
  - Discovery Channel store locations?
  - Toys and products sold by Discovery Channel stores?
  - Products based on the Animal Planet program?

Introduction

There can be multiple tasks for the same interpretation:

- **Query:** Carnegie Mellon University
  - Find the home page of CMU (Navigational)
  - Find the location of CMU (Location)
  - Check recent news about CMU (News)
  - Find pictures of CMU (Pictures)

Introduction: Why Diversification?

Relevance based ranking model's target:
- Maximize the expected user satisfaction in each position
- Mostly based on textual similarity with the original query

This leads to a ranking that
- Mainly focuses on the most probable intent
- Is good for some users, but bad for the other users

Diversification is a trade-off between robustness and relevance
- Reduce relevance for the most probable intent
- Increase robustness by covering multiple intents

When it is clear what this user wants, no need to diversify
- But user intent is often unclear

7

7

Introduction: Multiple Dimensions of Diversification

There are multiple levels of diversity
- Interpretation of the original query
  - What does this query refer to?
  - What aspects of this query does the user want?
- Different tasks for each interpretation
  - What task does the user want to complete?
    - » Image, Map, Video, Navigational, etc.

We mainly focus on diversification of interpretations
- But Google may treat the two as equally important

8

8

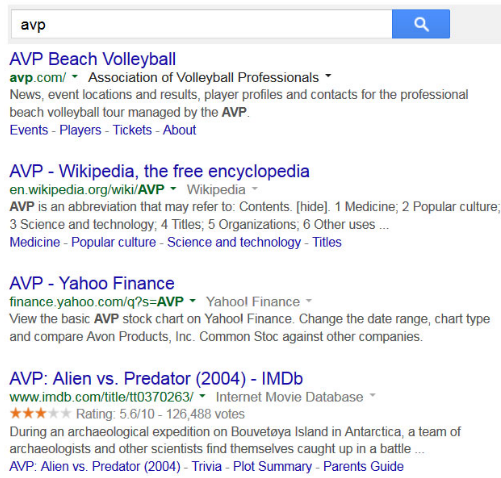
Introduction: Diversification of Interpretations

Query: avp

Interpretations:

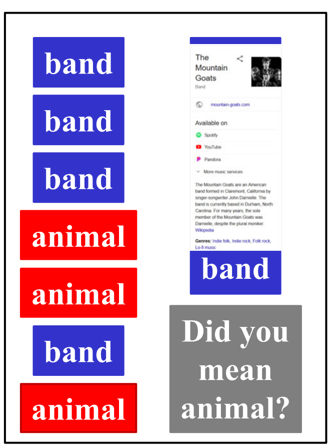
- Volleyball
- Wikipedia disambiguation page
- Company (Avon Products)
- Movie (Alien vs. Predator)

Purpose: Reduce the risk of focusing on wrong interpretations

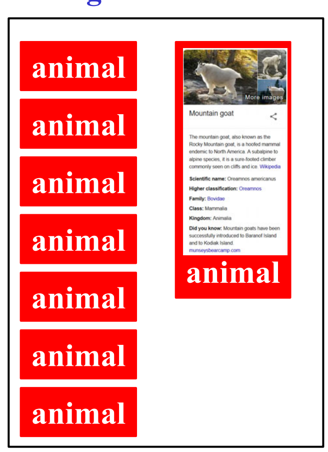


Introduction: Diversification of Interpretations

Query: mountain goats



After click on disambiguation dialog to select animal



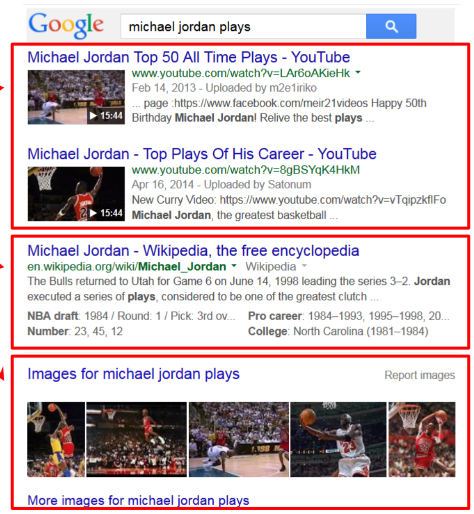
Introduction: Diversification for Tasks

Query: michael jordan plays

Tasks:

- Videos
- Information
- Images

Purpose: Make it easier to accomplish different tasks



Introduction

Usually diversification is done by re-ranking

- Use the query to produce an initial ranking R
  - Use your favorite / most effective method
- Re-rank the top N documents to produce a diverse ranking S

Query: avp

Initial Ranking R	Diverse Ranking S
volleyball	volleyball
volleyball	movie
movie	company
volleyball	volleyball

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
  - DSPApprox: Query intents (subtopics) discovery revisited

13

13

Diversity Metrics: P@k

Remember Precision at rank k (P@k)

- Easy to compute, easy to understand
- P@k ignores the diversity of the ranking

Ranking
Rel
Rel
NotRel
Rel
Rel **P@5 = 80%**
:
:

14

14

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

15

15

Diversity Metrics: Precision-IA@k

Intent-aware Precision at rank k

- Assume a query q has n intents $\{q_1, \ldots, q_n\}$
- Each intent may have a different probability $p(q_i \mid q)$
  - Typically in research publications $p(q_i \mid q)$ is uniform (1/n)
  - However, uniform $p(q_i \mid q)$ is not required
- Use some method to rank the documents for query q
- Calculate $P@k_{q_i}$ for each intent q_i
  - Use relevance of the document to query intent q_i
- Average the $P@k_{q_i}$ for each intent to produce Precision-IA@k

$$\text{Precision-IA}@k = \sum_{q_i} p(q_i \mid q) P@k_{q_i}$$

(Agrawal, et al., 2009)

16

16

Diversity Metrics: Precision-IA@k

Example: Precision-IA@5 for a query with <u>two</u> intents

$$\text{Precision-IA}@k = \sum_{q_i} p(q_i \mid q) P@k_{q_i}$$

$p(q_i|q) = 0.5$

Ranking 1	Ranking 2	Ranking 3
Rel_1	Rel_1	Rel_1
Rel_1	Rel_2	$\text{Rel}_{1,2}$
$\text{NotRel}_{1,2}$	$\text{NotRel}_{1,2}$	$\text{NotRel}_{1,2}$
Rel_1	Rel_2	Rel_2
Rel_1	Rel_1	Rel_1
Precision-IA@k	Precision-IA@k	Precision-IA@k
= ½×0.8 + ½×0	= ½×0.4 + ½×0.4	= ½×0.6 + ½×0.4
= 0.4 + 0	= 0.2 + 0.2	= 0.3 + 0.2
= 0.4	= 0.4	= 0.5

Relevant to intent 1

Uniform intent weights

Diversity Metrics: Precision-IA@k

Usually a good strategy is to give higher priority to documents that satisfy multiple intents

- That improved Precision-IA@k for Ranking 3 in the previous example
- Satisfying multiple intents with one document is an efficient way to lower risk
- We will see that idea again throughout the lecture

Diversity Metrics:
NDCG

Remember NDCG

- It considers the position of each document
- It allows multi-valued relevance assessments
- It ignores the diversity of the ranking

$$NDCG@k = Z_k \sum_{i=1}^{k} \frac{2^{R(i)} - 1}{\log_2(1+i)}$$

← **Gain (based on relevance)**
← **Discount (based on rank)**

19

19

Diversity Metrics:
α-NDCG

NDCG consists of two components

- Gain from a document worth R(k) at rank k: $G@k = 2^{R(k)} - 1$
- Discount for selecting a document at rank k: $D@k = \frac{1}{log_2(1+k)}$

α-NDCG uses an intent-aware gain calculation

$$G@k = \sum_{q_i} R(d_k, q_i) \boxed{(1-\alpha)^{r_{q_i,k-1}}}$$

Discount for covering an intent again

$R(d_k, q_i)$: Relevance of d_k to q_i

$$r_{q_i,k-1} = \sum_{j=1}^{k-1} R(d_j, q_i)$$

$$r_{q_i,0} = 0$$

Intent redundancy of d_k: Relevance of docs already selected for intent q_i

(Clarke, et al., 2008)

20

20

Diversity Metrics: α-NDCG

The gain vector for α-NDCG discounts intent redundancy

- 1-α controls the discount
- If α=0.5

  Value (j+1$_{th}$ reldoc, q_i) = 0.5 ×Value (j$_{th}$ reldoc, q_i)

  (each document is worth ½ the previous document)

(Clarke, et al., 2008)

21

21

Diversity Metrics: α-NDCG

The gain vector for α-NDCG discounts intent redundancy

- Assume α=0.5 and relevance values are {0, 1}

Ranking 1	G@k	Ranking 2	G@k	Ranking 3	G@k
Rel_1	1.000	Rel_1	1.000	Rel_1	1.000
Rel_1	0.500	Rel_2	1.000	$Rel_{1,2}$	1.500
$NotRel_{1,2}$	0.000	$NotRel_{1,2}$	0.000	$NotRel_{1,2}$	0.000
Rel_1	0.250	Rel_2	0.500	Rel_2	0.500
Rel_1	0.125	Rel_1	0.500	Rel_1	0.250

Rel: 1
NotRel: 0

22

22

Diversity Metrics: α-NDCG

The gain vector for α-NDCG discounts intent redundancy

- Assume α=0.5 and relevance values are {0, 1, 2}

Ranking 1	G@k	Ranking 2	G@k	Ranking 3	G@k
Rel_1	1.000	Rel_1	1.000	HRel_1	2.000
HRel_1	1.000	HRel_2	2.000	$\text{Rel}_{1,2}$	1.250
$\text{NotRel}_{1,2}$	0.000	$\text{NotRel}_{1,2}$	0.000	$\text{NotRel}_{1,2}$	0.000
HRel_1	0.250	Rel_2	0.250	Rel_2	0.500
Rel_1	0.031	Rel_1	0.500	Rel_1	0.125

HRel: 2
Rel: 1
NotRel: 0

Diversity Metrics

Both metrics are used

- E.g., the TREC 2009-2012 Diversity task
- E.g., All the recent diversification research papers

There are also many other evaluation metrics

- All of them have somewhat similar motivation and behavior

Which metric better matches your expectations?

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
  – DSPApprox: Query intents (subtopics) discovery revisited

Introduction: Types of Diversification Algorithms

Implicit

- Query intents are implicit in the document ranking
- Similar documents are assumed to cover similar intents

Explicit

- Query intents are specified explicitly
- Rerank documents so that all query intents are covered

Maximum Marginal Relevance (MMR)

Maximum Marginal Relevance (MMR) was one of the first diversification algorithms

(Interesting trivia: A SIGIR Test of Time award winner)

- Use the query to retrieve a ranking R of documents
- Use MMR to rerank the top N documents (build a new ranking S)
  - Select the 1st document based on how well it satisfies the query
  - Select subsequent documents based on two criteria
    - » How well it satisfies the query
    - » How different it is from documents ranked above it

(Carbonell and Goldstein, 1998)

27

27

Maximum Marginal Relevance (MMR)

MMR is a greedy algorithm

- At each step, select a document to append to the ranking

$$MMR \stackrel{def}{=} \arg\max_{d_i \in R \setminus S} (\lambda\, Sim(q, d_i) - (1-\lambda) \max_{d_j \in S} Sim(d_i, d_j))$$

Similar to the query **Different from the documents ranked above**

R: The initial ranking

S: Documents already selected for the diverse ranking

– Documents ranked above this position

$Sim(d_i, d_j)$: Use your favorite similarity metric

– E.g., vector space model or Jensen-Shannon Divergence

(Carbonell and Goldstein, 1998)

28

28

Maximum Marginal Relevance (MMR): Example with λ=0.6

Initial Ranking

d_1, 0.80
d_2, 0.78
d_3, 0.76
d_4, 0.74
d_5, 0.72
d_6, 0.70
:

Maximum Marginal Relevance (MMR): Example with λ=0.6

Initial Ranking

d_1, 0.80
d_2, 0.78
d_3, 0.76
d_4, 0.74
d_5, 0.72
d_6, 0.70
:

Diversified Ranking, Step 1

d_1

Maximum Marginal Relevance (MMR): Example with λ=0.6

Initial Ranking	Similarity to $\{d_1\}$	Diversified Ranking, Step 1
d_1, 0.80		d_1
d_2, 0.78	0.7	
d_3, 0.76	0.4	
d_4, 0.74	0.7	
d_5, 0.72	0.2	
d_6, 0.70	0.4	
:	:	

Maximum Marginal Relevance (MMR): Example with λ=0.6

Initial Ranking	Similarity to $\{d_1\}$	Score of d_i	Diversified Ranking, Step 1
d_1, 0.80			d_1
d_2, 0.78	0.7	$0.6 \times 0.78 - 0.4 \times 0.7 = 0.1880$	
d_3, 0.76	0.4	$0.6 \times 0.76 - 0.4 \times 0.4 = 0.2960$	
d_4, 0.74	0.7	$0.6 \times 0.74 - 0.4 \times 0.7 = 0.1640$	
d_5, 0.72	0.2	$0.6 \times 0.72 - 0.4 \times 0.2 = 0.3520$	
d_6, 0.70	0.4	$0.6 \times 0.70 - 0.4 \times 0.4 = 0.2600$	
:	:	:	

Maximum Marginal Relevance (MMR): Example with λ=0.6

Initial Ranking	Diversified Ranking, Step 2
d_1, 0.80	d_1
d_2, 0.78	d_5
d_3, 0.76	
d_4, 0.74	
d_5, 0.72	
d_6, 0.70	
:	

Maximum Marginal Relevance (MMR): Example with λ=0.6

Initial Ranking	ArgMax Similarity to $\{d_1, d_5\}$	Diversified Ranking, Step 2
d_1, 0.80		d_1
d_2, 0.78	0.7	d_5
d_3, 0.76	0.5	
d_4, 0.74	0.7	
d_5, 0.72		
d_6, 0.70	0.6	
:	:	

Maximum Marginal Relevance (MMR): Example with λ=0.6

Initial Ranking	ArgMax Similarity to $\{d_1, d_5\}$	Score of d_i	Diversified Ranking, Step 2
d_1, 0.80			d_1
d_2, 0.78	0.7	$0.6 \times 0.78 - 0.4 \times 0.7 = 0.1880$	d_5
d_3, 0.76	0.5	$0.6 \times 0.76 - 0.4 \times 0.5 = 0.2560$	
d_4, 0.74	0.7	$0.6 \times 0.74 - 0.4 \times 0.7 = 0.1640$	
d_5, 0.72			
d_6, 0.70	0.6	$0.6 \times 0.70 - 0.4 \times 0.6 = 0.1800$	
:	:	:	

35

35

Maximum Marginal Relevance (MMR): Example with λ=0.6

Initial Ranking	Diversified Ranking, Step 3
d_1, 0.80	d_1
d_2, 0.78	d_5
d_3, 0.76	d_3
d_4, 0.74	
d_5, 0.72	
d_6, 0.70	
:	

36

36

Maximum Marginal Relevance (MMR): Example with λ=0.6

Initial Ranking	ArgMax Similarity to $\{d_1, d_5, d_3\}$	Diversified Ranking, Step 3
d_1, 0.80		d_1
d_2, 0.78	0.8	d_5
d_3, 0.76		d_3
d_4, 0.74	0.7	
d_5, 0.72		
d_6, 0.70	0.6	
:	:	

Maximum Marginal Relevance (MMR): Example with λ=0.6

Initial Ranking	ArgMax Similarity to $\{d_1, d_5, d_3\}$	Score of d_i	Diversified Ranking, Step 3
d_1, 0.80			d_1
d_2, 0.78	0.8	$0.6 \times 0.78 - 0.4 \times 0.8 = 0.1480$	d_5
d_3, 0.76			d_3
d_4, 0.74	0.7	$0.6 \times 0.74 - 0.4 \times 0.7 = 0.1640$	
d_5, 0.72			
d_6, 0.70	0.6	$0.6 \times 0.70 - 0.4 \times 0.6 = 0.1800$	
:	:	:	

Maximum Marginal Relevance (MMR): Example with λ=0.6

Initial Ranking	Diversified Ranking, Step 4
d_1, 0.80	d_1
d_2, 0.78	d_5
d_3, 0.76	d_3
d_4, 0.74	d_6
d_5, 0.72	
d_6, 0.70	
:	

Maximum Marginal Relevance (MMR): Example with λ=0.6

Initial Ranking	ArgMax Similarity to $\{d_1, d_5, d_3, d_6\}$	Diversified Ranking, Step 4
d_1, 0.80		d_1
d_2, 0.78	0.8	d_5
d_3, 0.76		d_3
d_4, 0.74	0.9	d_6
d_5, 0.72		
d_6, 0.70		
:	:	

Maximum Marginal Relevance (MMR): Example with λ=0.6

Initial Ranking	ArgMax Similarity to $\{d_1, d_5, d_3, d_6\}$	Score of d_i	Diversified Ranking, Step 4
d_1, 0.80			d_1
d_2, 0.78	0.8	$0.6 \times 0.78 - 0.4 \times 0.8 = 0.1480$	d_5
d_3, 0.76			d_3
d_4, 0.74	0.9	$0.6 \times 0.74 - 0.4 \times 0.9 = 0.0840$	d_6
d_5, 0.72			
d_6, 0.70			
:	:	:	

41

41

Maximum Marginal Relevance (MMR): Example with λ=0.6

Initial Ranking	Diversified Ranking, Step 5
d_1, 0.80	d_1
d_2, 0.78	d_5
d_3, 0.76	d_3
d_4, 0.74	d_6
d_5, 0.72	d_2
d_6, 0.70	
:	

42

42

Maximum Marginal Relevance (MMR): Example with λ=0.6

Initial Ranking

d_1, 0.80
d_2, 0.78
d_3, 0.76
d_4, 0.74
d_5, 0.72
d_6, 0.70
:

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

Diversified Ranking, Step 5

d_1
d_5
d_3
d_6
d_2

43

43

Anything-to-Anything Similarity

The vector space can compare the similarity of any two vectors

- E.g., lnc.ltc
- This is convenient for a wide variety of tasks, e.g., clustering

BM25 is not designed for anything-to-anything similarity

Can it be done within the language modeling framework?

44

44

Probabilistic Similarity

There are many ways of measuring the similarity of two probability distributions

- Kullback-Leibler divergence is a popular measure

$$KL(d_1 \| d_2) = \sum_{t \in V} p(t \mid d_1) \ln \frac{p(t \mid d_1)}{p(t \mid d_2)}$$

$$p(t \mid d) = \frac{tf_{t,d}}{length(d)}$$

- But … KL divergence isn't symmetric

$KL(d_1 \| d_2) \neq KL(d_2 \| d_1)$

Term	d_1	d_2
apple	0.00	0.00
buy	0.02	0.00
camera	0.05	0.03
dog	0.00	0.00
image	0.07	0.02
like	0.03	0.00
mode	0.04	0.00
movie	0.03	0.00
up	0.01	0.00
…	…	…
zooms	0.02	0.05
Total		

45

45

Probabilistic Similarity

Jensen-Shannon Divergence is a symmetric and smoothed version of KL divergence

$$JS(x \| y) = \frac{1}{2} KL(x \| M) + \frac{1}{2} KL(y \| M)$$

$$M = \frac{x+y}{2}$$

This is simpler than it may look

$$JS(d_1 \| d_2) = \frac{1}{2} \sum_{t \in V} p(t \mid d_1) \log \frac{p(t \mid d_1)}{p(t \mid d_1 \cup d_2)} + p(t \mid d_2) \log \frac{p(t \mid d_2)}{p(t \mid d_1 \cup d_2)}$$

Often used for anything-to-anything similarity tasks

46

46

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
  - DSPApprox: Query intents (subtopics) discovery revisited

LeToR for Diversification

MMR is the diversified version of unsupervised retrieval

- Standard retrieval model + Similarity with previous documents
  - E.g., vector space model, language model
- Simple and intuitive
- No supervision used

Can we do it in learning to rank framework?

- Multiple sources of evidence are available
  - Relevance evidence
  - Document similarity evidence
- Many LeToR models are available for combining evidence

LeToR for Diversification: Directly Apply LeToR Models

Approach: Directly fit a standard learning-to-rank model to optimize a diversity measure (e.g., α-NDCG)

- Use same model and features that we saw in the LeToR lecture
- Use diversification based training data
  - E.g., pointwise, pairwise, or listwise training data
- Let machine learning algorithms figure out how to produce diversified rankings

This approach does not work well

- No consistent gains on various metrics
- No consistent performances on training data vs testing data

(Santos, et al., 2011)

49

49

LeToR for Diversification: Directly Apply LeToR Models

Approach: Directly fit a standard learning-to-rank model to optimize a diversity measure (e.g., α-NDCG)

Why doesn't this approach work well?

- There are no diversification features
  - Nothing about document-to-document similarity
  - Nothing about sub-intents
- Relevance-based features can not produce a diverse ranking
- Thus, the diversified training data just confuses learning models

(Santos, et al., 2011)

50

50

LeToR for Diversification: Relational-LeToR

Diversification requires relevance <u>and</u> relationship features

- Traditional LeToR features to handle relevance
- Relationships among documents to handle diversity

This is <u>implicit</u> diversification, so the relationship is <u>similarity</u>

- Similarities to higher-ranked documents (as done by MMR)
- Multiple ways to measure such similarity
- Combined by LeToR Model

The <u>relational</u> version of LeToR can be viewed as a supervised version of MMR

(Zhu, et al. 2014)

51

51

Relational-LeToR: Similarities between Documents

Similarity/relation measures between documents:

- Text similarity:
  – E.g. KL-Divergence of document language models
- Topic similarity:
  – E.g. Cosine between docs' topic distributions in topic models
- Category similarity:
  – E.g. Overlap of docs' categories in a predefined ontology
- Link similarity:
  – E.g. whether docs have hyperlinks to each other
- URL similarity:
  – E.g. whether docs are from same website or not.

(Zhu, et al. 2014)

52

52

Relational-LeToR: Features from Similarities

Combine similarity with all higher-ranked documents by:

- Minimal similarity with previous documents
- Average similarity with previous documents
- Maximal similarity with previous documents

Each similarity measure is combined to one feature

- The relationship of the current doc to all higher-ranked documents
  - In one similarity dimension
- E.g. Maximum text similarity with higher-ranked docs

Each combination is one method in experiments

- Like a hyper-parameter to choose (by cross-validation)

(Zhu, et al. 2014)

53

53

Relational-ListMLE: An Example Model

ListMLE can be easily modified to handle relational features

- Other LeToR models can be modified too
- ListMLE is the first one studied, perhaps also the easiest

Recall the sequential assumption of ListMLE

- Documents are picked one by one from top to bottom
  - Same with MMR's sequential assumption
- Each position is independent of previous ones
  - Relational features break this independence assumption

(Zhu, et al. 2014)

54

54

Relational-ListMLE: New Generative Process

Relational-ListMLE's new generative process

- The ranking is assembled iteratively from candidate documents
- Given a set of $S_i = \{d_1, \ldots, d_n\}$ of candidate documents
  - Pick the best candidate d_i from S_i to appear at rank i

$$p(d_i|S_i;w) = \frac{\exp(w_r \cdot x_i + w_d \cdot y_i)}{\sum_{x_j \in S_i} \exp(w_r \cdot x_j + w_d \cdot y_j)}$$

  - $S_{i+1} = S_i \setminus d_i$
  - Repeat until all documents are reranked

w_r: Relevance weights
x_i: Relevance features for d_i
w_d: Diversity weights
y_i: Diversity features for d_i

***$p(d_i|S_i;w)$* depends on higher-ranked documents via similarities**

(Zhu, et al. 2014)

55

55

Relational-ListMLE: Learning and Ranking

Learning is exactly the same as with ListMLE

- Write down likelihood, use MLE

In ranking, iteratively pick the doc with largest $p(d_i|S_i, w)$

- Cannot directly calculate ranking score as ListMLE does
  - Now documents depend on previous positions
- Complexity is $O(n^2)$
  - In each of n positions
    - » Pick the best document from remaining documents
    - » Update the similarity features of other documents (O(n))
  - n isn't huge because diversification is a re-ranking method
    - » n < 1000 in most cases

(Zhu, et al. 2014)

56

56

Relational-ListMLE: Performance

Web Track 2009, Diversity task data (similar results for 2010)

Method	ERR-IA		α-NDCG		NRBP	
QL	0.1637		0.2691		0.1382	
ListMLE	0.1913	(+16.86%)	0.3074	(+14.23%)	0.1681	(+21.64%)
MMR_{list}	0.2022	(+23.52%)	0.3083	(+14.57%)	0.1715	(+24.09%)
$xQuAD_{list}$	0.2316	(+41.48%)	0.3437	(+27.72%)	0.1956	(+41.53%)
$PM\text{-}2_{list}$	0.2294	(+40.13%)	0.3369	(+25.20%)	0.1788	(+29.38%)
SVMDIV	0.2408	(+47.10%)	0.3526	(+31.03%)	0.2073	(+50.00%)
$R\text{-}LTR_{min}$	**0.2714**	(+65.79%)	0.3915	(+45.48%)	**0.2339**	(+69.25%)
$R\text{-}LTR_{avg}$	0.2671	(+63.16%)	**0.3964**	(+47.31%)	0.2268	(+64.11%)
$R\text{-}LTR_{max}$	0.2683	(+63.90%)	0.3933	(+46.15%)	0.2281	(+65.05%)
TREC-Best	0.1922		0.3081		0.1617	

No diversity
Explicit diversification (given subtopics)
Relational-ListMLE

(Zhu, et al. 2014)

Implicit Methods: Discussion

Implicit method characteristics

- Doesn't require prior knowledge about possible query intents
  - Thus called "implicit"
- Selects documents that match the query well
- Produces a diverse ranking
- Doesn't favor any query intent
  - Popular intents don't get more attention than rare intents
  - Is this good or bad?

MMR is a classic unsupervised method

R-LeToR is a newer supervised method

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
  - DSPApprox: Query intents (subtopics) discovery revisited

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.