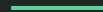


SLWWW at the NTCIR-15 WWW-3 Task

Masaki Muraoka
Zhaohao Zeng
Tetsuya Sakai

Motivations



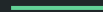
Aim

- Reproduce THU's WWW-2 run.^[1]
 - THUIR-E-CO-MAN-Base-2 (LambdaMART)
- Comparison the performance of BM25 and Learning-to-rank models.
 - *can the gain of LambdaMART over BM25 as reported by THUIR (Tsinghua University) at WWW-2 be successfully replicated and reproduced?*^[2]

[1] Yukun Zheng, Zhumin Chu, Xiangsheng Li, Jiaxin Mao, Yiqun Liu, Min Zhang, and Shaoping Ma. 2019. THUIR at the NTCIR-14 WWW-2 Task. In Proceedings of NTCIR-14. 472–480.

[2] Tetsuya Sakai, Sijie Tao, Zhaohao Zeng, Yukun Zheng, Jiaxin Mao, Zhumin Chu, Yiqun Liu, Maria Maistro, Zhicheng Dou, Nicola Ferro, and Ian Soboroff. 2020. Overview of the NTCIR-15 We Want Web with CENTRE (WWW-3) Task. In Proceedings of NTCIR-15. to appear.

Methodology



Dataset

- Training Dataset
 - LETOR4.0 ^[3]
- Validation Dataset
 - NTCIR-13 WWW English Test Collection
- Test Dataset
 - Rerank the provided baseline documents

[3] Tao Qin and Tie-Yan Liu. 2013. Introducing LETOR 4.0 Datasets. (2013). arXiv:arXiv:1306.2597

Features

- Extracted four fields.
 - Title
 - Anchor text
 - URL
 - Whole document
- 8 features for each field.
- idf for the whole document is reused for
 - Title
 - Anchor text
 - URL
- Features for the training set are provided.

Table 1: Extracted features

feature name
tf (term frequency)
idf (inverse document frequency)
tf*idf
document length
BM25 score
LMIR.ABS
LMIR.JM
LMIR.DIR

Learning to Rank models

- Ranklib library ^[4]
- 8 learning to rank models
 - MART
 - RankNet
 - RankBoost*
 - AdaRank*
 - Coordinate Ascent*
 - LambdaMART*
 - ListNet
 - Random Forests

Table 3: Mean nDCG@10 scores over the WWW-1 topic set (n=100) based on the WWW-2 CENTRE qrels

Algorithm	nDCG@10
MART	0.3510
RankNet	0.3313
RankBoost	0.3746
AdaRank	0.3699
Coordinate Ascent	0.3470
LambdaMART	0.2763
ListNet	0.2865
Random Forests	0.3184

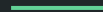
* Selected for use in WWW-3 runs.

[4] Van Dang. 2012. The Lemur Project-Wiki-RankLib. Lemur Project. <http://sourceforge.net/p/lemur/wiki/RankLib>

BM25

- Tuned on the WWW-2 test collection.
 - Grid search maximizing nDCG@10.
 - $k = 1.1$, $b = 0.7$

Results



Results

- nDCG, ERR : BM25, AdaRank >> LambdaMART
- nDCG : Baseline >> LambdaMART

Run	nDCG	Q	ERR	iRBU
RankBoost	0.5189	0.5366	0.6397	0.8773
AdaRank	0.6227	0.6359	0.7345	0.9040
Coordinate Ascent	0.5719	0.5822	0.7061	0.8919
LambdaMART	0.4465	0.4531	0.5804	0.8220
BM25	0.6291	0.6445	0.7362	0.8981

Discussion



Poor performance of LambdaMART

- Performed statistically significantly worse than the baseline.
- Could not fully replicated the methodology.
 - a. Used features provided by the LETOR4.0 dataset.
 - b. Used the idf of the whole document.

Poorly Performing Topics

Table 5: Hardest topics in terms of Average nERR over all WWW-3 runs

ID	Average nERR	Content
0132	0.1354	Friends
0169	0.1447	The origin of paper
0144	0.2715	Origin of Teacher's Day
0147	0.2872	bird nest effect
0153	0.3219	freeweblayouts

"Friends"

- Description:
 - "Friends is a very famous TV series, you want to know in which year it first aired"
- All runs used the content field exclusively.
 - Ambiguous content field.
 - Retrieved documents containing "friends" used in different contexts.
- 18 relevant documents in qrels.
 - Six contains "1994"; all with relevance grade 2.
 - Only one got relevance grade of 3.

BACK

132. T1: Friends D: Friends is a very famous TV series, you want to know in which year it first aired.

PREV

NEXT

- clueweb12-1715wb-71-15707
- clueweb12-0104wb-68-03262
- clueweb12-0010wb-28-12649
- clueweb12-1810wb-42-07117
- clueweb12-1106wb-33-07246
- clueweb12-0012wb-55-20420
- clueweb12-0901wb-75-23115
- clueweb12-1813wb-43-33492
- clueweb12-0206wb-96-30292
- clueweb12-0102wb-34-05933
- clueweb12-0300wb-07-24245
- clueweb12-1300tw-51-00390
- clueweb12-1616wb-59-14010

Topics A B C D E F G H I J K L M N O P Q R S T U V W

💡 *Did you know:* "One is apt to overestimate beauty when it is rare" Mark Twain

Popular Topics

Attitude quotes
Being In Love With Your Best Friend quotes
Being Single quotes
Broken Friendship quotes
Cute Missing You quotes
Disappointment quotes
Ex Boyfriend quotes
Facebook Status quotes
Facts quotes
Fake Friends quotes
Funny Facebook Status quotes
Happy Friday quotes
Heart Touching quotes
I Love You For Him quotes
It Is What It Is quotes
Letting Go quotes
Life quotes
Love quotes
Memories quotes
Missing Someone quotes
Moving On quotes
Relationships quotes
Someone Special quotes

Fake Friends Quotes & Sayings

Related Results

Fake Friends Quotes About: [Broken Heart](#) [Life Lesson](#) [Advice](#) [Fake People](#) [Fake Smile](#) [True Friend](#) [Childhood](#) [Being Fake](#) [Backstabbing Friend](#) [Fake Friendship](#) [Two Faced People](#) [Being Two Faced](#) [Be Yourself](#) [Being Yourself](#) [Fake Girls](#) [Facebook Status](#) [Problems](#) [Enemies](#) [Hoes](#) [Funny Facebook Status](#) [Rumors](#) [Two Faced Friends](#) [Cheating Boyfriend](#) [Being Used By People](#) [Being Used](#) [True Friends](#) [Backstabber](#) [Backstabbing](#) [Cocky](#) [Shit Talking](#) [Being Real](#) [I Hate You](#) [Fuck Friends](#) [Clocks](#) [Drama](#) [Games](#) [Miss You](#) [Dont Trust Nobody](#) [Fake Girl](#)

Sorted by: [New Fake Friends Quotes](#) | Most Popular Total: 164 Quotes

Broken heart again. Another lesson learned. Better know your friends. Or you will get burned. 52 up, 14 down

Unknown quotes

About: [Broken Heart quotes](#) [Life Lesson quotes](#) [Fake Friends quotes](#) [Advice quotes](#)

[Like 171](#) [Tweet](#)

Fake friends are autumn leaves; they are scattered everywhere. 50 up, 6 down

Unknown quotes

About: [Fake Friends quotes](#) [Fake People quotes](#)

[Like 205](#) [Tweet](#)

Fake friends believe your fake smile, true friends see the tears your hiding inside. 96 up, 11 down

NONREL: 0 REL: 0 H.REL: 0 ERROR: 0

Relevance: 3

Comparison of Rankers

- We examine the ERR scores of two rankers:
 - AdaRank
 - BM25
- Why did our learning-to-rank model did not outperform BM25?

Table 6: Topics where BM25 scored substantially better than AdaRank (nERR)

ID	Content	BM25	AdaRank	Diff
0160	akron beacon journal	0.9994	0.3034	0.6960
0132	Friends	0.5635	0.1602	0.4033

"akron beacon journal"

- Description
 - You want to find the official website of Akron Beacon Journal
- Average topic nERR of ~ 0.79
- Mean nERR (AdaRank and BM25) of ~ 0.73

Table 6: Topics where BM25 scored substantially better than AdaRank (nERR)

ID	Content	BM25	AdaRank	Diff
0160	akron beacon journal	0.9994	0.3034	0.6960
0132	Friends	0.5635	0.1602	0.4033

AdaRank run results

Rank	DocID	Rel	URL
1	clueweb12-1312wb-14-19400	0	www.sportspyder.com
2	clueweb12-1310wb-87-17847	0	www.sportspyder.com
3	clueweb12-0109wb-53-29184	1	www.granatino.com
4	clueweb12-0400tw-75-09336	2	www.howtomakeyourcathappy.com
5	clueweb12-0305wb-90-14750	3	ohio.kaango.com
6	clueweb12-0105wb-90-17385	3	www.config.com
7	clueweb12-0305wb-81-03188	2	ohio.kaango.com
8	clueweb12-0303wb-79-23517	2	ohio.kaango.com
9	clueweb12-0304wb-93-33348	3	ohio.kaango.com
10	clueweb12-0304wb-66-18382	2	ohio.kaango.com

* URLs in red redirect to the akron beacon homepage

BM25 run results

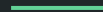
Rank	DocID	Rel	URL
1	clueweb12-0105wb-90-17385	3	www.config.com
2	clueweb12-0810wb-12-03213	3	www.afsc.net
3	clueweb12-0904wb-72-21384	3	the330.com
4	clueweb12-0904wb-87-20357	3	the330.com
5	clueweb12-1300tw-48-15018	2	screamradius.com
6	clueweb12-0303wb-79-23517	2	ohio.kaango.com
7	clueweb12-0915wb-38-16574	2	www.knightcomm.org
8	clueweb12-0400tw-75-09336	2	www.howtomakeyourcathappy.com
9	clueweb12-0304wb-93-33348	3	ohio.kaango.com
10	clueweb12-0206wb-96-02296	2	www.ohio.com

* URLs in red redirect to the akron beacon homepage

Comparison

- Both rankers have 5 fully relevant documents, albeit a different ranking.
- Inconsistent gradings?
 - Documents from the same domain receiving different scores.
 - Less relevant (or irrelevant) documents receiving higher scores.

Conclusion



Conclusion

- Unable to reproduce THU's results in WWW-2.
- Our Learning-to-rank models do not outperform BM25.
- Can and should rankers utilize the description field?
- Quality of relevance assessment needs to be explored.

Thank you

Q/A
