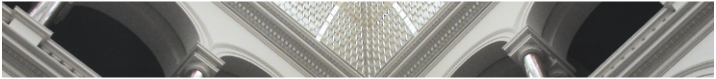
 

Defining a Meaningful Baseline for News Recommender Systems

Benjamin Kille and Andreas Lommatzsch | Technische Universität Berlin



Motivation

Baseline Candidates

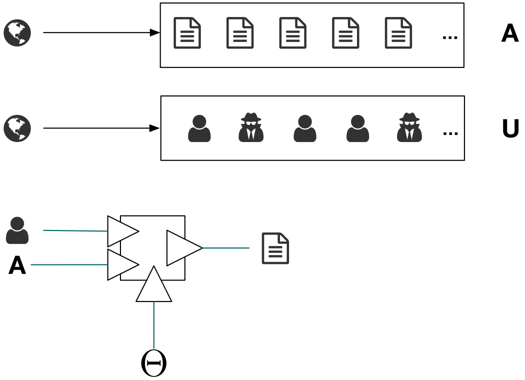
Experiment

Results

Conclusion

Motivation

What do **News Recommender Systems** do?



Motivation

How to tell which algorithm to use?

1. Define a way to measure how well the algorithm does
2. Define a method to compare against
3. Conduct experiments
4. Analyse results

What baseline to choose? (2.)

Motivation

Reference	Baseline
(Das et al. 2007)	popularity
(Garcin et al. 2014)	popularity
(Gao et al. 2011)	term-frequency
(L. Zheng et al. 2013)	term-frequency
(Cantador, Castells, and Bellogín 2011)	keywords
(Okura et al. 2017)	keywords
(Lihong Li et al. 2010)	ϵ-greedy bandit
(Lei Li, L. Zheng, et al. 2014)	CF, CBF
(Lu et al. 2015)	CF, CBF
(Lei Li and T. Li 2013)	(Das et al. 2007), (Liu, Dolan, and Pedersen 2010), (Chu and Park 2009), (Lei Li, L. Zheng, et al. 2014), (Lei Li, D. Wang, et al. 2011)
(G. Zheng et al. 2018)	(Rendle 2010), (Cheng et al. 2016), (Lihong Li et al. 2010), (Huazheng Wang, Wu, and Hongning Wang 2016)
(Hongwei Wang et al. 2018)	(Rendle 2010), (J. Wang et al. 2017), (Huang et al. 2013), (Cheng et al. 2016), (Guo et al. 2017), (Covington, Adams, and Sargin 2016), (Xue et al. 2017)
(Khattar et al. 2018)	(Rendle et al. 2009), (He, Zhang, et al. 2016), (Kumar et al. 2017), (Musto et al. 2016), (He, Liao, et al. 2017)

Motivation

Desiderata for Baselines:

- simple to implement
- competitive $\rightarrow$ **expressiveness**
- compatible with available data

Random

Rational

Optimise the coverage!

Method

Recommend a random item from the collection

Parameter

$\theta = \Delta T \rightarrow A = A_{t_{\mathrm{now}} - \Delta T}$

Recency

Rational

Readers always prefer to read the latest news!

Method

Consider the time of publication and recommend the most recent item

Popularity

Rational

What is interesting to many will most likely fit individuals!

Method

Record how frequently readers engage with each article and recommend the article read most often

Parameter

$\theta = \Delta T \rightarrow A = A_{t_{\mathrm{now}} - \Delta T}$

Content-based Filtering

Rational

Reader shows interesting in a topic, so lets give them more of the same!

Method

Each item has a topic/category assigned. Consider the current item and find the most popular item in the category

Collaborative Filtering

Rational

Collaborative Filtering works always well!

Method

(Das et al. 2007) (MinHash)

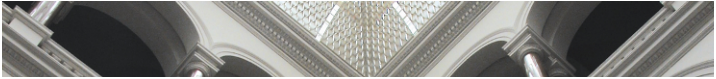
Reading Sequences

Rational

Hmm ... I have seen that many readers read *A* first and then *B*; lets recommend *B* whenever someone is currently looking at *A*

Method

1. for all sessions with at least two articles collect all Sequences
2. given the current article determine the most likely successor



Technische Universität Berlin

Circular Buffer

Rational

Capture what is going on!

Method

1. define a circular list
2. as readers engage with articles add them to the list
3. recommend the last item added to the list which is different from the current item

Parameter

Θ = L (length of the circular list)

Trends

Rational

What is trending should be a good recommendation, should it not?

Method

1. bin the number of reads for each article and hour
2. compute the trend for each article
3. recommend the article with the steepest ascent

Parameter

$\Theta = \Delta T$ (number of data points/length of time window)

Experimental Design

For all sessions $s \in S$

for all events $x \in s$

recommend an article $a \in A$

if x has been previously recommended count as success $y(s) = y(s) + 1$

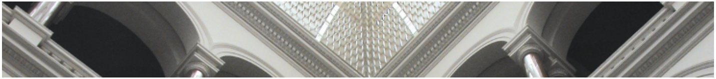
Compute evaluation metrics:

$$R_y = |S|^{-1} \sum_{s \in S} y(s) \tag{1}$$

$$R_s = |S|^{-1} \sum_{s \in S} 1_{\{y(s)>0\}} \tag{2}$$

Data Set

Statistic	Publisher A	Publisher B	Publisher C
Number of sessions	17 019 523	22 683 047	54 272 242
Number of reads	36 859 823	175 930 128	105 998 109
Events per session	2.17	7.76	1.95
Number of sessions with 1 read	10 529 390	416 506	34 998 380
Proportion of session with 1 read	61.19 %	1.84 %	64.49 %



Results

Baseline	Publisher A $R_y \cdot 10^{-5}$	Publisher A $R_s \cdot 10^{-5}$	Publisher B $R_y \cdot 10^{-5}$	Publisher B $R_s \cdot 10^{-5}$	Publisher C $R_y \cdot 10^{-5}$	Publisher C $R_s \cdot 10^{-5}$
random	6.26	6.21	6.16	6.11	1.73	1.69
random (6h)	87.69	85.59	64.30	62.43	13.72	13.45
random (12h)	56.02	55.05	43.05	42.22	9.99	9.81
random (24h)	37.75	37.29	32.69	32.21	8.06	7.88
random (48h)	27.06	26.87	23.80	23.47	6.11	6.04
popular	934.30	653.56	431.17	82.23	2025.96	1581.96
popular (6h)	1009.02	722.28	282.33	53.83	2027.88	1582.75
popular (12h)	1039.16	755.82	526.83	100.99	2031.92	1585.04
popular (24h)	1089.32	774.56	1166.36	216.55	2037.30	1587.64
popular (48h)	1140.12	781.17	1388.02	258.87	2043.20	1590.65
recency	708.23	647.72	1017.66	200.90	38.17	34.36
sequences	**18 542.49**	**14 650.12**	0.14	0.03	1523.25	1256.07

Results (continued)

Baseline	Publisher A $R_y \cdot 10^{-5}$	Publisher A $R_s \cdot 10^{-5}$	Publisher B $R_y \cdot 10^{-5}$	Publisher B $R_s \cdot 10^{-5}$	Publisher C $R_y \cdot 10^{-5}$	Publisher C $R_s \cdot 10^{-5}$
circular buffer (100)	8069.65	6583.47	**11 747.88**	**3049.56**	2069.87	1583.18
circular buffer (200)	8069.65	6583.47	11 748.00	3049.56	2074.45	1585.22
circular buffer (500)	8069.65	6583.47	11 748.00	3049.56	**2074.76**	**1585.35**
circular buffer (1000)	8067.65	6583.47	11 748.00	3049.56	**2074.76**	**1585.35**
content-based	330.10	310.96	286.01	89.93	4.27	4.22
collaborative filtering	5864.39	4861.47	4648.87	845.90	132.68	116.79
trends (2h)	8599.78	6833.47	11 237.24	2066.37	1608.16	1246.95
trends (6h)	6437.37	4743.85	6398.07	1173.67	1784.49	1381.72
trends (12h)	5805.29	4205.00	4481.68	823.05	1795.34	1392.33
trends (24h)	4858.50	3418.07	1580.30	292.20	1723.92	1344.13

Conclusion

- Circular Buffer and Trending perform well on all publishers
- some Baselines exhibit high variance between publishers
- Random and CBF appear insufficiently competitive

Future Work

- Consider additional baseline candidates (suggestions welcome)
- Explore time in space complexity with detailed measurements
- Analyse variance over time to check if baselines' performance fluctuates

Questions?