

KDD Cup 2016

李朋軒

2016/03/17

KDD Cup 2016

<https://kddcup2016.azurewebsites.net/>

Background

- To find influential nodes in a social network
- To measure the impact of research institutions
- Whose papers are accepted the most

Task

- To predict the rank of 741 institutions
- By their number of accepted papers in 2016
- Target conferences:
 - SIGIR, SIGMOD, SIGCOMM (March 28)
 - KDD, ICML (April 28)
 - FSE, MobiCom, MM (May 28)

(\$10,000 at stake)

Evaluation

- DCG
 - Input: ratings, ranking
 - Output: score
- NDCG
 - *R: ground truth ratings*
 - $\frac{DCG(R, \text{predicted ranking})}{DCG(R, \text{ground truth ranking})}$

Data

<http://research.microsoft.com/en-us/projects/mag/>

<https://scholar.google.com>

<http://dl.acm.org/>

Past publications

- Microsoft Academic Graph
- 3677 full papers
 - from selected conferences
 - from 2011 to 2015

Paper-Author-Institution

- Microsoft Academic Graph
- ACM Digital Library
 - Selenium

Citations

- HTTP REQUEST google scholar
- Captcha
 - Firefox cookie

Model

<http://scikit-learn.org/>

Instance

- An institution an instance
- Feature vector
 - Conferences: 8,
 - Years: 3,
 - Accepted papers, and
 - Their citations
- Label
 - Accepted papers

Model

- Ridge regression with cross-validation
- Train on 2014
- Validate on 2015

Illustration

Accepted Papers 2014

```
26.0 stanford university
24.5 ibm
19.6 microsoft
19.5 university of texas at austin
16.6 hong kong university of science and technology
13.2 university of southern california
12.8 purdue university
12.7 university of washington
11.7 peking university
10.0 university of amsterdam
8.4 university of california santa barbara
8.0 university of waterloo
7.4 rutgers university
7.4 national university of singapore
7.0 university of massachusetts amherst
6.9 massachusetts institute of technology
6.8 tsinghua university
6.8 university of pennsylvania
6.7 nanyang technological university
6.5 university of toronto
2014 idcg=103.531
```

Accepted Papers 2015

```
19.0 stanford university
15.1 ibm
13.0 microsoft
11.6 university of southern california
11.5 university of texas at austin
8.8 peking university
8.6 nanyang technological university
8.0 hong kong university of science and technology
7.9 university of technology sydney
7.8 university of california santa barbara
7.8 university of wisconsin madison
7.8 the chinese university of hong kong
7.7 university of washington
7.6 university of california berkeley
7.5 university of amsterdam
7.2 university of toronto
6.5 university at buffalo
6.1 carnegie mellon university
6.0 university of waterloo
5.9 national university of singapore
2015 idcg=75.2152
```

Accepted Papers

```
117.8 ibm
111.1 stanford university
87.3 microsoft
57.3 university of washington
57.3 university of texas at austin
53.1 hong kong university of science and technology
45.9 university of southern california
43.2 university of toronto
41.9 peking university
41.8 purdue university
36.9 duke university
32.9 the chinese university of hong kong
32.8 university of california santa barbara
32.2 university of amsterdam
29.0 university of wisconsin madison
27.8 national university of singapore
27.7 massachusetts institute of technology
27.1 university of illinois at urbana champaign
27.1 rutgers university
25.0 nanyang technological university
```


Train on all conferences

	NDCG 2014	NDCG 2015
All	0.96	0.92
ICML	0.96	0.66
KDD	0.96	0.63
SIGMOD	0.96	0.53
SIGIR	0.96	0.35
SIGCOMM	0.96	0.50
FSE	0.96	0.47

Train on each conference

	NDCG 2014	NDCG 2015
All	0.96	0.92
ICML	0.95	0.82
KDD	0.87	0.73
SIGMOD	0.80	0.72
SIGIR	0.92	0.63
SIGCOMM	0.86	0.59
FSE	0.70	0.56

SIGIR Leader Board

Rank	Name	Score
1	ΣΑ··^Σ SPN	0.9880314204
2	Yabanai?	0.9810058037
2	FEG&NSSOL@DataVeraci	0.9810058037
2	MakeAWish	0.9810058037
2	thesis	0.9810058037
6	T310B	0.9750032986
7	NLPR_IVA_XYZ	0.9741013026
8	ecustwm	0.9583897792
9	NTU	0.956714601
10	CA AdTech Seminar	0.9552574841