Automatically Assessing Wikipedia Article Quality by Exploiting Article–Editor Networks

Xinyi Li, Jintao Tang, Ting Wang, Zhunchen Luo and Maarten de Rijke

University of Amsterdam, National University of Defense Technology, China Defense Science and Technology Information Center

Introduction

- **What?**

  We study the quality assessment of Wikipedia articles.

- **Why?**

  Wikipedia articles vary in quality and only a minority of them are manually evaluated high quality articles. Since manually labeling articles is inefficient, it is essential to automatically assess article quality.

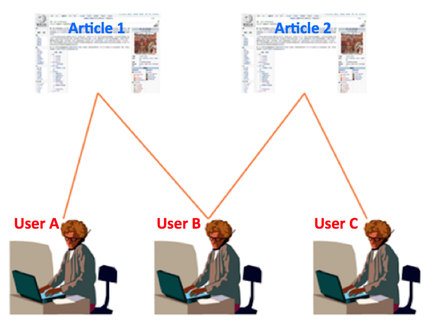
470 million articles in English

0.1% are featured(best quality ones)

30,000 active editors per month...and dropping

- **How?**

  We view this task as a ranking problem by exploiting the article-editor network. We combine existing manual evaluations on Wikipedia as features for automatic ranking.



Models

We have developed several models for estimating Wikipedia article quality based on the article-editor network:

- **Pagerank**
  –Treat both articles and editors as nodes connected by edges that represent editing relations.

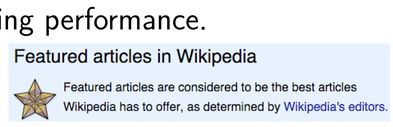
- **Simple Weighted (SW) model**
- **Complex Weighted (CW) model.**
  –Consider weights between edges.

- **Simple Weighted Probabilistic (SWP) model**
- **Complex Weighted Probabilistic (CWP) model**
  –Incorporate manual evaluation results and assign an article's initial value as its probability of being high quality.

Evaluation

- **Assessing article quality by ranking**

  Using **featured articles** (the best quality articles on Wikipedia) as the gold standard to measure ranking performance.



- **Metric**

  1.Recall scores at the first N items in the result set.
  2.Precision-recall curves.

Dataset

Table: **Statistics of datasets**.

Category	#articles	#editors	#featured articles
Chemistry	7,796	392,055	36
Meteorology	4,218	187,637	138
Geography	38,543	1,360,508	180

Experiment 1

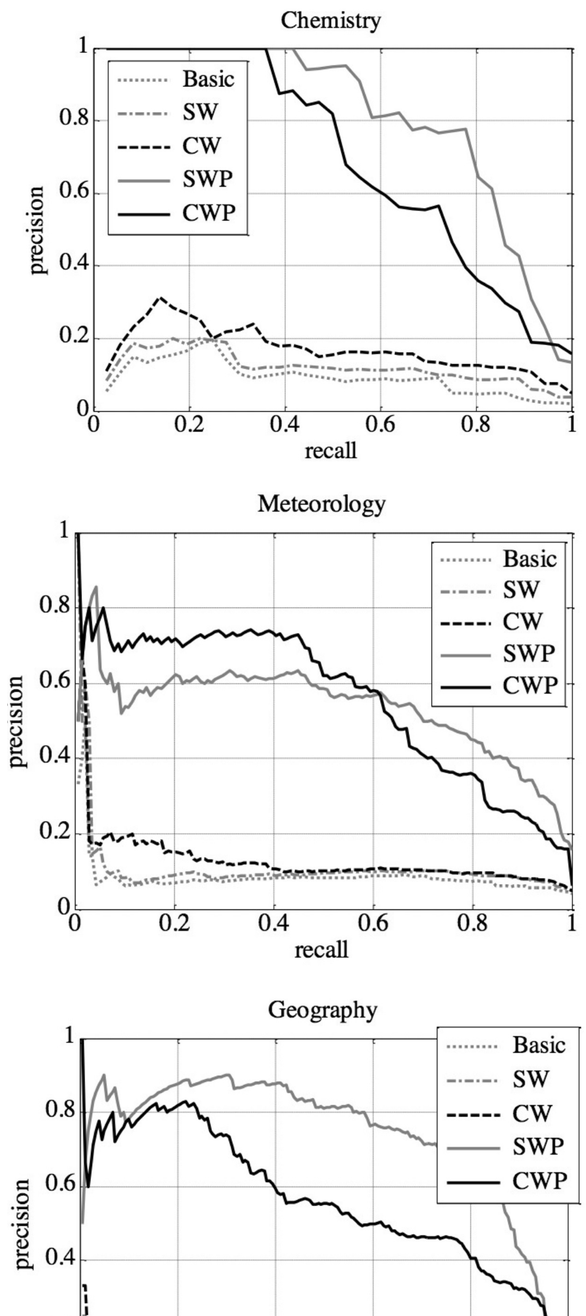
Aim: examine the impact of the number of featured articles on ranking performance

Recall (N) of SWP and CWP in different categories

featured%	r@100		r@200		r@300		r@400	
	SWP	CWP	SWP	CWP	SWP	CWP	SWP	CWP
chemistry								
25%	.556	.363	.767	.667	.867	.793	.440	.874
50%	.644	.378	.778	.694	.861	.833	.972	.883
75%	.756	.400	.911	.744	.956	.911	1.000	.944
meteorology								
25%	.111	.092	.246	.175	.365	.317	.498	.421
50%	.101	.103	.274	.165	.438	.346	.607	.486
75%	.140	.114	.346	.200	.517	.357	.703	.514
geography								
25%	.173	.086	.342	.168	.426	.283	.496	.369
50%	.163	.069	.357	.182	.497	.317	.562	.422
75%	.149	.051	.376	.162	.518	.327	.596	.407

Experiment 2

Aim: compare model performance when using all featured articles for initialization



Precision-recall curves for the baseline (Basic), simple weighted (SW), complex weighted (CW), simple weighted probabilistic (SWP), complex weighted probabilistic (CWP) model.

Conclusion

- Link structure is valuable for ranking in this setting.
- Weighted models perform better than basic models.
- Combination of existing manual evaluations with the article-editor network yields a state-of-the-art solution for assessing article quality.

UNIVERSITY OF AMSTERDAM