

Automatically Assessing Wikipedia Article Quality by Exploiting Article–Editor Networks

Xinyi Li^{1,2}, Jintao Tang¹, Ting Wang¹, Zhunchen Luo³, and Maarten de Rijke²

¹ National University of Defense Technology, Changsha, China
{tangjintao, tingwang}@nudt.edu.cn

² University of Amsterdam, Amsterdam, The Netherlands
{x.li, derijke}@uva.nl

³ China Defense Science and Technology Information Center, Beijing, China
zhunchenluo@gmail.com

Abstract. We consider the problem of automatically assessing Wikipedia article quality. We develop several models to rank articles by using the editing relations between articles and editors. First, we create a basic model by modeling the article-editor network. Then we design measures of an editor’s contribution and build weighted models that improve the ranking performance. Finally, we use a combination of featured article information and the weighted models to obtain the best performance. We find that using manual evaluation to assist automatic evaluation is a viable solution for the article quality assessment task on Wikipedia.

1 Introduction

Wikipedia is the largest online encyclopedia built by crowdsourcing, on which everyone is able to create and edit the contents. Its articles vary in quality and only a minority of them are manually evaluated high quality articles.¹ Since manually labeling articles is inefficient, it is essential to automatically assess article quality. Content quality criteria are known to help retrieval; in a web setting they are often based on link structure [7, 8] but in the setting of social media and collaboratively created content, content-based features are often used [11]. Here, we study the quality assessment of Wikipedia articles by exploiting the article-editor network. We view this task as a ranking problem. Our task is motivated by the assumption that automatic procedures for assessing Wikipedia article quality can help information retrieval that utilizes Wikipedia resources [2] and information extraction on Wikipedia [13] to obtain high quality information.

There have been different approaches to the content quality assessment problem. One branch of research uses simple metrics, such as article length, number of links and citations etc. [1, 6, 9, 12]. These authors do not consider the interactions between editors and articles, which differentiates Wikipedia from traditional encyclopedias. Other work takes into account the network of articles and editors. Hu et al. [4] proposes what they call a probabilistic review model to rank articles. The model is tested on a dataset of only 242 articles. Suzuki and Yoshikawa [10] uses a combination of survival ratio method and link analysis to score articles. They use relative evaluation metrics to measure the performance of models. It remains to be seen to which degree they can achieve satisfactory ranking results in more realistic settings.

¹ Only 0.1% of all Wikipedia articles are featured articles.

We examine the editing actions of editors and find that the majority of them are field-specific, i.e., they specialize in a certain category of articles. These field-specific editors outnumber all-around editors to a great extent. Since the editor-article networks of different categories only share very few nodes, ranking articles should be done in separate categories. As featured articles are manually-tagged high quality articles, we select them as the ground truth for our task. We develop several models to rank articles by quality. Our first motivation is to see if the importance of a node in the network can indicate quality. So we develop a basic PageRank-based model. Additionally, instead of treating links as equal in the basic model, we tweak the model by putting weights on the links to reflect the difference of editor contributions. Finally, we utilize existing manual evaluation results to improve automatic evaluation. So we incorporate manual evaluation results into our model. We use articles of different quality levels to measure the levels of editors, and then assist ranking.

The experiments carried out on multiple datasets covering different fields show that ranking performance is related to the number of high quality articles we utilize. In particular, the higher the percentage of high quality articles used, the better the ranking performance. We also find that the basic model does not yield satisfactory ranking results, but that using weights boosts performance.

2 Models

We introduce the models and explain how each model is computed, including a baseline model, weighted models, and weighted models with probabilistic initial value.

2.1 Baseline Model

First, we develop a basic quality model based on Pagerank. PageRank is widely applied for ranking web pages, where pages are seen as nodes and hyperlinks as edges [7]. The node value represents its importance in the network. In our basic model we treat both articles and editors as nodes connected by edges that represent editing relations. For instance, if article A is edited by B then there is a bidirectional edge that connects A and B. The value of the nodes are distributed through the edges during each iteration of the PageRank computation. As shown in (1), the value of node v is determined by nodes in the set $U(v)$ that connect to it, where $N(u)$ is the number of edges that point out of node u .

$$PR(v) = (1 - d) + d \sum_{u \in U(v)} \frac{PR(u)}{N(u)}. \quad (1)$$

In this basic model, we give all nodes the same initial value and iteratively compute the node value until they converge. The articles will then be ranked by node value.

2.2 Weighted Models

The baseline model treats edges as equal. However, consider an article that has multiple editors, which is quite common. When the value of the article node is distributed toward its editors during computation, editors that make a higher contribution should get more.