A Multi-site Collaborative Sampling for Web Accessibility Evaluation

Zhi Yu\textsuperscript{1,2,3(✉)}, Jiajun Bu\textsuperscript{1,2,3}, Chao Shen\textsuperscript{1,2,3}, Wei Wang\textsuperscript{1,2,3}, Lianjun Dai\textsuperscript{4}, Qin Zhou\textsuperscript{4}, and Chuanwu Zhao\textsuperscript{5}

\textsuperscript{1} Zhejiang Provincial Key Laboratory of Service Robot, College of Computer Science, Zhejiang University, Hangzhou, China
\{yuzhirenzhe, bjj, shenchao, wangwei_eagle\}@zju.edu.cn
\textsuperscript{2} Alibaba-Zhejiang University Joint Institute of Frontier Technologies, Hangzhou, China
\textsuperscript{3} Ningbo Research Institute, Zhejiang University, Ningbo, China
\textsuperscript{4} China Disabled Persons’ Federation Information Center, Beijing, China
\{dailianjun, zhouqin\}@cdpf.org.cn
\textsuperscript{5} Zhejiang Toman Intelligent Manufacturing Technology Co., Ltd, Xinchang, China
zhaocw@zjtoman.com

Abstract. Many sampling methods have been used for web accessibility evaluation. However, due to the difficulty of web page feature extraction and the lack of unsupervised clustering algorithm, the result is not very good. How to optimize the manual workload of different websites under the premise of ensuring that the overall manual workload remains the same during multi-site collaborative sampling is an important issue at present. To resolve the above problems, we propose a multi-site collaborative sampling method to obtain the final sampling result of each website. The effectiveness of the two sampling methods proposed in this paper is proved by experiments on real website datasets.

Keywords: Information accessibility · Accessibility evaluation · Sampling method · Active learning · Semi-supervised clustering

1 Introduction

The rise of various e-commerce websites has connected our online and offline lives. However, most of these websites were not designed with information access methods for people with disabilities in mind. For example, the visually impaired can only read the content of a web page by reading screen software and traversing the focus. If the website is designed without considering the compatibility of the reading screen software, it will block the people with visual impairment from accessing the information. Instead, according to the survey in 2010, there were 85.02 million disabled people in China, and the number of disabled people in the world exceeded 1 billion. In order to meet the requirements of information accessibility, anyone can obtain information...
equally and effectively under any circumstances. The information gap of disabled people in the current Internet information age deserves attention.

In order to enable people with disabilities to enjoy the benefit of Internet information on an equal basis with able-bodied people, the World Wide Web Consortium (W3C) published the first website content accessibility guide [1] (WCAG 1.0) in 1995. Website provides effective guidance. With the wide application of this guide in countries around the world, and considering some existing limitations, the W3C released WCAG2.0 in November 2008 [2]. WCAG2.0 has made a lot of supplements to WCAG1.0. It is designed with practicality in mind. It is no longer limited to specific technologies, and has made many compatibility considerations for future technologies.

Based on this, we have developed a web accessibility evaluation system for the Chinese government. Through accessibility assessment of the website, we obtain website detection results and suggestions to guide website accessibility reconstruct. However, affected by the lack of automatic detection tools, the detection process still requires the participation of accessible experts. For the thousands of web pages, manual inspection of each webpage becomes impractical, so web page sampling came into being. Because the quality of the sampling effect will directly affect the website’s ultimate barrier-free overlap, which sampling algorithm is currently a major problem.

Existing web accessibility evaluation sampling methods give more consideration to how to extract which pages within a web site as a sample collection for accessibility assessment. In actual testing tasks, a batch of websites are usually evaluated at the same time. However, due to the limitation of accessibility testing expert resources, how to reasonably allocate the number of expert resources to each website (that is, how to determine when the total number of samples is fixed, the sample and quantity of each website) is often overlooked. The current number of samples for each website is usually determined by the number of pages on the website. If the number of pages on the website is larger, the sample set obtained by sampling will be larger, so more accessibility testing experts will be assigned for testing, and vice versa, The less. However, this method of allocating expert resources according to the number of website pages does not take into account the impact of website complexity, resulting in a higher overall sampling error.

In the process of multi-site collaborative sampling, it is unreasonable to only consider the number of website pages to determine the sample and quantity of each website. We should also fully consider the impact of differences in website structure. To solve the above problem, this paper proposes Multi-site collaborative sampling. The effectiveness of this method proposed in this paper is proved by experiments on real website datasets.

2 Related Work

In this part, we briefly review the related works in web page sampling methods for Web accessibility evaluation.

Random sampling guarantees that every page in the website will be sampled with equal probability. However, the distribution of checkpoints is not evenly distributed. For example, a verification code is a checkpoint that appears on only a few pages. For a
website with a large number of pages, when random sampling is used, there will be a high probability that web pages containing verification code will not be extracted, causing sampling errors.

The Ad-hoc sampling method proposed by W3C/WAI and UWEM [3] first sets sampling rules in advance by domain experts. However, the shortcomings of this method are also obvious. Due to the complex website structure and the huge number of websites, it is difficult to accurately find different types of web pages. In addition, due to human intervention, this method requires high cost and has certain subjectivity.

The random walk sampling method [4] consists of two phases: 1) the walk phase: starting from the homepage of the website, selecting a webpage from the URL out of the webpage with probability $p$ to visit; returning to the previous visit pages with probability $1-p$; 2) Sampling phase: Take a specified number of web pages from the pages visited during the wandering phase as the final sample. Because the random walk sampling method does not need to obtain a complete set of web pages, compared with random sampling, the calculation and storage overhead are relatively small. In addition, compared with Ad-hoc sampling, because no manual intervention is required, this method is more objective. However, this method tends to select lower-level web pages as samples, ignoring deep-level web pages, so that the most representative web pages in the entire website cannot usually be obtained.

In order to make the sampling samples reflect the characteristics of checkpoints, King et al. Proposed a sampling method based on the distribution of checkpoints [5]. This method first uses automatic detection tools to count the automatic detection results of each web page, and then based on the automatic detection of these web pages. The detection results are clustered. Finally, according to the cluster size, random sampling is performed from each cluster to form the final sample set. This method is based on the inaccurate assumption that the distribution of automatic checkpoints is the same as the distribution of manual checkpoints, which leads to errors in the detection of manual detection items.

Zhang [6] proposed a semi-supervised method based on active prediction to try to solve the problem of web page sampling and evaluation based on web structure checkpoints. This method again actively learns appointments to the picked webpages, and actively selects the best alternative webpages for manual detection, so that it can get smaller errors while reducing its detection cost. The limitation of this method is that it divides the accessibility results into only two categories: pass and fail, this simple pass and fail cannot judge the performance of the two websites on the checkpoint more accurately.

### 3 Sampling Method

How to optimize the manual workload of different websites under the premise of ensuring that the overall manual workload remains the same during multi-site collaborative sampling is an important issue at present.

Inspired by the idea of turning accessibility sampling problems into active prediction problems proposed by [6], we use an active learning method based on a sample selection strategy that combines uncertainty and representativeness. Compared with an
active learning method based on a statistical model, this method has a lower computational complexity, but the difficulty of this method lies in how to select each iteration.

We use Hinted SVM [7] to ensure that the training query hyperplane can better distinguish untrained sample web pages. By integrating the Active Learning Webpage Sampling Model with Hinted SVM, we can obtain the active learning webpage sampling algorithm AL-HSVM (Active Learning Hinted SVM) based on the hinted support vector machine.

Initially, we randomly select a small number of web pages Pl for manual detection, and then in each iterative process of active learning, first select a subset Ph from the undetected web page set Pu as the hint pool, and then use Hinted SVM training from Pl and Ph. The training obtained the hyperplane h. Finding the closest webpages to h from the undetected webpage set Pu can be regarded as the webpage with the highest uncertainty and the best representation, that is to say, it needs manual detection in the next cycle. And all the similar pages will be removed from the hint pool.

In the process of collaborative sampling of multiple websites, because the numbers of pages corresponding to each website are different, we find the most uncertain pages according to Hinted SVM in each website, and then delete them in the hint pool. The number of similar pages is also different. For a website with a simple structure, it only contains a small number of page templates, but the number of pages generated by these page templates is large. When we delete similar pages, we can usually delete a large number of similar pages. Therefore, the size of the current hint pool can reflect the complexity of the website structure to a certain extent. If in the current iteration step, the hint pool still contains a lot of web pages, it means that there are still many different templates in this site that have not been detected by the query hyperplane, so more detection resources need to be allocated for it to make hint pool size reduction. In order to avoid the impact of the difference in the number of pages on different websites, we use the ratio of the number of pages in the hint pool to the total number of pages on the website to horizontally compare the size of the hint pool for different websites.

Therefore, our sampling strategy is: First, each website randomly selects a small number of web pages for detection, obtains a set of detected web pages, and initializes the hint pool for all undetected web pages. In each subsequent iteration, we calculate the hint pool size of each site, select the site with the largest size and perform an AL-HSVM iteration on it. While selecting web pages for expert detection, all pages with a similar structure to this page are removed from the hint pool. Repeat the above steps until the number of samples reaches the specified total number of samples.

4 Experiment

We do an experiment to verify the validity of our method in real Web accessibility evaluation data. We will start with dataset description.

This experimental data set was collected by the accessibility detection system on October 24, 2018, and these sites were also tested for accessibility based on the detection entries and accessibility scores for each web page were obtained. There are five websites in this experimental data set, the first two of which are provincial
disability federations, the third one is a website of a unit directly under the China Disabled Persons’ Federation, and the last two are social websites with a high daily frequency.

4.1 Experiment for Single Website Sampling

First we do the experiment to verify the prediction model based on the web pages selected by active learning can obtain better prediction results with less detection cost than the prediction model established by randomly selected web pages.

As can be seen from the Fig. 1, which shows the accuracy of different sampling methods, AL-HSVM not only considered the uncertainty of the detection result of the webpage when selecting the webpage, but also considered the representativeness of the webpage in the entire website. Therefore, compared to AL-Uncertainty, a sampling method that only considers the uncertainty of webpage detection results, and AL-Representative sampling method, which only considers the representativeness of webpages, AL-HSVM is significantly higher than them in prediction accuracy. It can be found that the AL-HSVM method can obtain higher prediction accuracy under fewer web pages.

![Fig. 1. The accuracy of different sampling method.](image)

4.2 Experiment for Multi-site Collaborative Sampling

Another experiment is mainly used to verify the effectiveness of multi-site collaborative sampling algorithm based on active learning. It is verified that the method can more reasonably allocate the sample number of each website and determine which pages are extracted from each website under a certain total sample number, so that the total sampling error is minimized.
In addition to the collaborative learning algorithm based on active learning proposed in this article, the comparison algorithms in this experiment include:

In this experiment, we use the sum of the sampling errors of all websites as the evaluation criteria. The result is shown in Fig. 2.

![Fig. 2. The accuracy of different sampling method.](image)

We can clearly see that the total sampling error of the multi-site collaborative sampling algorithm based on active learning is significantly lower than the other two methods. Furthermore, let us observe that when the total number of samples is 75, the number of samples and the sampling error representation of each website.

5 Conclusion

We propose a multi-site collaborative sampling method based on active learning. This method aims to more reasonably determine the number of samples for each website and which pages to use as samples, under the condition of manual detection costs, so as to minimize the overall sampling errors. This method uses a hinting-based support vector machine’s active learning sampling algorithm. When selecting the next webpage to be detected, in addition to combining the uncertainty of the webpage detection results, it also treats the undetected webpages as hint, check the distribution of web pages as one of the basis for selection. After that, through the improved hint pool reduction strategy and the multi-site collaborative sample selection strategy, each time a sample is taken from the website with the smallest hint pool to obtain the final sampling results of each website. At the same time, many different types of experiments are designed. The sampling method proposed in this paper is compared with the current commonly used accessibility evaluation sampling methods. The performance of each sampling method on multiple different types of website data sets is verified.
Although the sampling method proposed in this paper effectively reduces the sampling error, there are still some shortcomings. In the multi-site collaborative sampling method, we divide the detection results of each web page into two types: Pass and Fail. However, the result is too high. It is rough and cannot effectively reflect the accessibility differences between different web pages. In the future, we can try to refine the accessibility detection results.

Acknowledgments. This work is supported by Alibaba-Zhejiang University Joint Institute of Frontier Technologies, The National Key R&D Program of China (No. 2018YFC2002603, 2018YFB1403202), Zhejiang Provincial Natural Science Foundation of China (No. LZ13F020001), the National Natural Science Foundation of China (No. 61972349, 61173185, 61173186) and the National Key Technology R&D Program of China (No. 2012BAI34B01, 2014BAK15B02).

References

1. Chisholm, W., Vanderheiden, G., Jacobs, I.: Web content accessibility guidelines 1.0. Interactions 8(4), 35–54 (2001)
2. World Wide Web Consortium: Web content accessibility guidelines (WCAG) 2.0 (2008)
3. Velleman, E., Velasco, C.A., Snaprud, M., et al.: D-WAB4 unified web evaluation methodology (UWEM 1.0). Technical report, WAB Cluster (2006)
4. Ulltveit-Moe, N., Snaprud, M., Nietzio, A., et al.: Early results from automatic accessibility benchmarking of public European web sites from the european internet accessibility observatory (EIAO) (2006)
5. King, M., Thatcher, J.W., Bronstad, P.M., et al.: Managing usability for people with disabilities in a large web presence. IBM Syst. J. 44(3), 519–535 (2005)
6. Zhang, M., Wang, C., Yu, Z., et al.: Active learning for Web accessibility evaluation. In: Proceedings of the 14th Web for All Conference on the Future of Accessible Work, p. 16. ACM (2017)
7. Li, C.L., Ferng, C.S., Lin, H.T.: Active learning using hint information. Neural Comput. 27(8), 1738–1765 (2015)