Generating Heatmap for Unknown Documents towards Readability Measurement

Jayasankar Santhosh
University of Kaiserslautern &
German Research Center for
Artificial Intelligence (DFKI)
Kaiserslautern, Germany
Jayasankar.Santhosh@dfki.de

Shoya Ishimaru
University of Kaiserslautern &
German Research Center for
Artificial Intelligence (DFKI)
Kaiserslautern, Germany
Shoya.Ishimaru@dfki.de

Andreas Dengel
University of Kaiserslautern &
German Research Center for
Artificial Intelligence (DFKI)
Kaiserslautern, Germany
Andreas.Dengel@dfki.de

ABSTRACT

The key idea behind this paper is to generate fixation heatmap of unknown documents to visualize and determine the focus areas in a document as a first step towards the readability measurement of the document. The data samples were collected by conducting experiment with nine participants reading 15 documents and the proposed method was to predict the fixation duration of each word in the documents. A Random Forest Regression model was used to predict the fixation duration per word and we achieved a mean regression score (R^2) of 0.757 for all the documents.

KEYWORDS

Eye tracking, fixation, heatmap, readability, regression

ACM Reference Format:
Jayasankar Santhosh, Shoya Ishimaru, and Andreas Dengel. 2020. Generating Heatmap for Unknown Documents towards Readability Measurement. In *25th International Conference on Intelligent User Interfaces Companion (IUI '20 Companion), March 17–20, 2020, Cagliari, Italy*. ACM, New York, NY, USA, 2 pages. https://doi.org/10.1145/3379336.3381495

1 INTRODUCTION

For many years, researchers have been investigating on methods to asses the quality of documents and identifying the user interest in the document content. The development of eye tracking technology has proved to be a vital advancement in determining the cognitive states like interest, focus, presence and consciousness of a user. An eye tracker is embedded with cameras, projectors and different algorithms that would capture the eye position and generate the eye movement patterns using infrared beams on the eyes. As investigated by Biedert *et al.*, utilizing eye tracking is a modern and objective way to quantify readability [1].

By using Eye tracking, it is possible to identify the user focus areas in a document. However, it is cumbersome and not realistic asking people to always mount a mobile eye tracker while reading. In order to tackle this bottleneck, we propose a method to visualize, machine learning based predicted eye movements over an unknown document using heatmap which helps in determining the reading pattern and the areas of a document where the user will probably spent more time in reading. This is a first step towards the readability measurement system, where the fixation duration of each word in a document is predicted using the document content features and utilizing the predicted word fixation duration along with the corresponding word coordinates, the fixation heatmap of a document is generated.

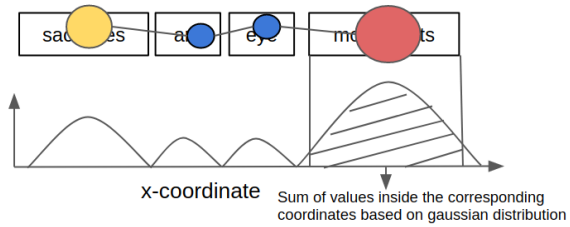

Figure 1: The fixation heatmap values corresponding to each word coordinate is summed to find the fixation heat value.

2 APPROACH

The eye movement pattern of various users were computed by using an eye tracker that could measure the fixation duration, saccade length, gaze X and Y coordinates providing information about reading pattern over the documents. The features from the document content were extracted and a Random Forest Regression model was used to predict the fixation duration per word for all the participants.

2.1 Ground Truth Recording

After detecting fixations and saccades from raw gaze recorded by an eye tracker, the heatmap of the documents were generated using kernel density estimation. Since a user reading a document does not mandatorily fixate on each and every word, the calculation of fixation duration per word becomes tough. In order to tackle this issue, each word is wrapped in a rectangular box based on coordinate values of the word and the corresponding fixation heatmap values inside each box will be summed up to get the fixation heat value per word as shown in Figure 1. We define the summed value as fixation heat value. Similar to x-coordinate values, the y-coordinate values are also estimated. The mean fixation heat value for each word in the document was computed by combining all the participant data and it was used as the labels for prediction.

IUI '20 Companion, March 17–20, 2020, Cagliari, Italy

ACM ISBN 978-1-4503-7513-9/20/03.
https://doi.org/10.1145/3379336.3381495

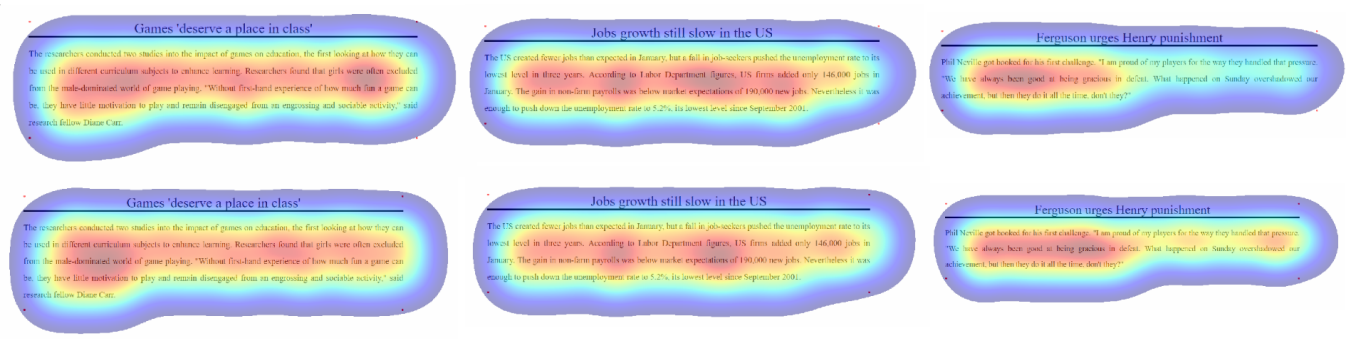

Figure 2: Three examples of our heatmap generation (top: ground truth, bottom: generated). The model for the heatmap generation was trained in a document-independent manner.

2.2 Feature Calculation

The following lexical and syntactical features from the document content were used for prediction. The length of each word in the document was used as the main feature. The other features were the log likelihood properties of the word like lexical, syntactical and total surprisal and ambiguity associated with each word [2]. These features were computed using an incremental probabilistic top down parser [4]. The syntactic features like the parts of speech tag (NN, VB, IN, JJ) was extracted for each word in the documents and was also used for prediction.

3 EXPERIMENTAL DESIGN

We used Tobii 4C eye tracker with a pro license key to track eye movements of the participants. The eye tracker was mounted on a desktop monitor and the documents were displayed on the monitor. Participants were asked to calibrate the eye tracker before reading each document in order to obtain the data without any errors.

Our experiment involved 10 Master pursuing students reading 15 documents each. All the documents were selected from BBC news articles and belonged to different categories like business, technical, politics, sports and entertainment. The reason behind the category wise selection was to ensure that documents varied in reading difficulty. In addition, we displayed a single paragraph step by step to record precise eye movements. The documents, when displayed had four dots as surrounding borders and this was done with the motive to limit the fixations within the document boundary. Before reading the documents, all the participants were requested to fixate on the four surrounding dots so as to ensure better accuracy in eye tracking. Due to the errors in eye tracking, one participant data was left out.

4 RESULTS AND ANALYSIS

A Random Forest Regression model was trained with *Leave-One-Document-Out* approach, where in each iteration one document will be used for testing and all other documents for training. A grid-search cross validation was performed to obtain the optimal hyper parameters. With the text features, we were able to achieve a mean regression score (R^2) of 0.757 and a normalized root mean squared error of 0.282. Using the Random Forest feature importance, it was observed that the word length was the most significant feature in the model prediction with the importance value of 0.972.

Using the predicted fixation heat values, the heatmap of each document was generated and main focus areas of the document were determined. Figure 2 depicts the groundtruth and corresponding predicted fixation heatmaps for three different documents and it can be observed that the model was able to predict the main focus areas (red colour region) as seen in the ground truth to an extent.

5 CONCLUSION AND FUTURE WORK

This paper focused on estimating the fixation duration of each word in an unknown document and generating the fixation heatmap based on predicted fixation heat values. The traditional, lexical and syntactical text features were computed from the document content and was used to train the model. The mean fixation heat value per word was predicted which was obtained by combining all the participant data. Using the predicted fixation heat values, the heatmap for each of the documents were generated.

The main future work after this study is to propose an assessment method of readability based on the fixation duration heat map. In addition, there are several potential ideas including personalizing the fixation heat map (i.e., providing parameters of language skill and knowledge level of the reader [3]) and generating lower-level behaviors including not only fixation durations but also fixation coordinates and saccades.

ACKNOWLEDGMENTS

This work is supported by JSPS KAKENHI (17K00276, 17K12728).

REFERENCES

[1] Ralf Biedert, Andreas Dengel, Mostafa Elshamy, and Georg Buscher. 2012. Towards robust gaze-based objective quality measures for text. In *Proceedings of the Symposium on Eye Tracking Research and Applications*. 201–204.
[2] John Hale. 2001. A probabilistic Earley parser as a psycholinguistic model. In *Proceedings of the second meeting of the North American Chapter of the Association for Computational Linguistics on Language technologies*. Association for Computational Linguistics, 1–8.
[3] Shoya Ishimaru, Syed Saqib Bukhari, Carina Heisel, Jochen Kuhn, and Andreas Dengel. 2016. Towards an intelligent textbook: eye gaze based attention extraction on materials for learning and instruction in physics. In *Proceedings of the 2016 ACM International Joint Conference on Pervasive and Ubiquitous Computing: Adjunct*. 1041–1045.
[4] Brian Roark. 2001. Probabilistic top-down parsing and language modeling. *Computational linguistics* 27, 2 (2001), 249–276.