Personalized Unknown Word Detection in Non-native Language Reading using Eye Gaze

Rui Hiraoka
Nara Institute of Science and Technology, Japan
hiraoka.rui.hj9@is.naist.jp

Hiroki Tanaka
Nara Institute of Science and Technology, Japan
hiroki-tan@is.naist.jp

Sakriani Sakti
Nara Institute of Science and Technology, Japan
ssakti@is.naist.jp

Graham Neubig
Nara Institute of Science and Technology, Japan*
neubig@is.naist.jp

Satoshi Nakamura
Nara Institute of Science and Technology, Japan
s-nakamura@is.naist.jp

ABSTRACT

This paper proposes a method to detect unknown words during natural reading of non-native language text by using eye-tracking features. A previous approach [5] utilizes gaze duration and word rarity features to perform this detection. However, while this system can be used by trained users, its performance is not sufficient during natural reading by untrained users. In this paper, we 1) apply support vector machines (SVM) [1] with novel eye movement features that were not considered in the previous work and 2) examine the effect of personalization. The experimental results demonstrate that learning using SVMs and proposed eye movement features improves detection performance as measured by F-measure and that personalization further improves results.

CCS Concepts

•Human-centered computing → Human computer interaction (HCI) ;

Keywords

Eye movement; unknown word detection; natural reading

1. INTRODUCTION

Unknown words appear more often in non-native language than native language, and thus unskilled non-native readers need to use dictionaries and machine-translation software to assist themselves in reading. However, the performance of machine translation between many languages is not often sufficient for usage in this situation, and manual search of dictionaries can be time-consuming. Our goal is to develop an automatic and quick unknown word detection system, allowing for more efficient and appropriate display of translations and explanations to help users reading non-native language. We focused on eye-gaze input to achieve rapid and effortless system. Eye tracking allows a user to read a document in their non-native language without using hands e.g. keyboard or mouse. Therefore, one of the benefits of use eye-gaze based interfaces is that they are faster than using a mouse, as many studies show [14][13]. Hyrskykari [5] has proposed a gaze-based unknown word detection approach that utilizes total gaze duration, word rarity, and a simple threshold function. After user training, their approach achieve high detection performance for unknown words (recall : 91.0%, false positive : 2.4%). However, without user training, it had significantly lower detection performance (recall : 28.8%, false positive : 0.9%). Therefore, it is not effective for untrained users. This paper proposes a new method for eye-gaze based unknown word detection during natural reading that aims to achieve acceptable performance even when used by untrained users. We extend Hyrskykari's approach [5] by using machine learning, defining a number of new features, and selecting the most effective ones. We also examined the effect of personalization (subject dependency). The experimental evaluation confirmed the effectiveness of 1) using a novel eye movement feature, max gaze duration, and 2) personalization to individual users.

2. RELATED WORK

This work focuses on how to detect unknown words using eye movement data while a reading non-native language. For approaches without eye movement, an unknown word detection method based on click logs was proposed by Ehara [3]. This approach uses a large click-log corpus to predict unknown words. In contrast, eye movement has been considered as a major indicator of human reading strategies. Rayner clarified the relation between eye movement and unknown words showing that the duration of fixation tends to be longer on low-frequency words and the number of regressive eye movements tends to increase. Kunze and Gomez proposed an algorithm to estimate language skills from eye movement [8][9]. This work showed fixation tends to occur frequently on difficult words. To the authors' knowledge, Hyrskykari has proposed the only gaze based method to automatically detect unknown words [5]. In this previous work, she developed iDict, which is an application to detect

*Now at Carnegie Mellon University, USA.

unknown words and display their explanation. This paper attempts to improve unknown words detection performance by using other eye movement features and personalization.

3. UNKNOWN WORD DETECTION

In this section, we explain Hyrskykari's work including a description of its features and threshold-based detection method.

3.1 Features

Eye movement indicators of human reading strategies and activities are fixation, saccade, and regression [10]. Fixations are a type of eye movement that consists of staring at a single location. Gazes are defined as the sum of all fixations on a certain region. Saccades are a quick, simultaneous movement of the right and left eyes between two or more phases of fixation. Regressions are backward eye movements. Text documents are formed by a sequence of N words, $W = \{w_1, w_2, ..., w_N\}$. A sequence of fixations $F = \{f_1, f_2, ..., f_M\}$ is given from recording of eye movement data. f_k is a coordinate on the target screen. Each feature is defined formally as follows:

gaze duration: $g(w_x)$
: is consecutive sum of all fixation durations which occurred on word w_x prior to saccade to another word.

total gaze duration on word w_x: $g_{\text{total}}(w_x)$
: is total sum of all fixation durations mapped to word w_x including regressive fixations.

rarity of word w_x: $rare(w_x)$
: For each word in the stimulus document, word rarity is calculated in a large corpus[1]. This is done by calculating the rank of word w_x's frequency in the corpus $rank(w_k)$ and rounding values less than 100 or more than 6000 up and down respectively:

$$rare(w_x) = \begin{cases} 100 & (rank(w_k) < 100) \\ 6000 & (rank(w_k) > 6000) \\ rank(w_k) & (\text{else}). \end{cases} \quad (1)$$

3.2 Threshold function

Previous work uses a threshold function for the detection

$$th(w_x) = th_h - \frac{th_h - th_l}{100 - 6000}(rare(w_x) - 100), \quad (2)$$

where, $th(w_x)$ is the total gaze duration threshold determining whether word w_x is unknown or known. th_h is the threshold of total gaze duration for high-frequency words ($rare(w_k) \leq 100$) in the corpus. th_l is the threshold total gaze duration for low-frequency words ($rare(w_k) \geq 6000$) in the corpus. th_h, th_l are defined as below.

$$th_h = \mu_h + a_h\sigma_h, \text{ and } th_l = \mu_l + a_l\sigma_l,$$

where μ_h and μ_l are the mean total gaze duration of high- and low-frequency words in the corpus. Likewise, σ_h and σ_l are standard deviations, and a_h, a_l are parameters to optimize detection performance to a given data set ($a_h > 0$, $a_l > 0$). The prediction is given whether $g_{\text{total}}(w_x)$ is larger than $th(w_x)$. We used this method as a baseline to compare with our proposed method.

[1]Specifically previous work used British National Corpus (2005) at http://www.natcorp.ox.ac.uk/

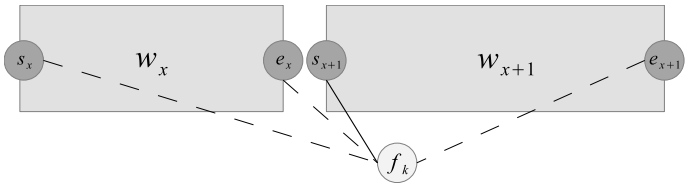

Figure 1: Fixation mapping. s_{x+1} is the nearest coordinate to f_k so f_k belongs to w_{x+1}. Therefore, $w(k) = x+1$ in this example.

4. PROPOSED METHOD

In this paper, we propose a method for unknown word detection that uses a classifier with new eye gaze features, and additionally personalize this to each individual user.

4.1 Fixation mapping

Fixations should be mapped to each word of the stimulius document. For each word, eye movement features are extracted based on the mapped fixations. These features include gaze duration, extracted pupil diameter size, and regressions. In general, eye tracking data is a sequence of gaze points sampled at a certain sampling rate, which are classified automatically into fixations and saccades by using software provided with the eye tracker.

To map fixations to the words in stimuli documents, we obtain the coordinates of w_x's start point s_x and end point e_x. An algorithm to determine the word fixation on which fixation f_k focuses is based on Euclid distance between fixation f_k and the xth word w_x's start point s_x and end point e_x respectively. Fixation f_k belongs to the word which has the minimum distance point out of all of the start and end points. Figure 1 shows an example of fixation mapping. The reason why we use the edges of each word instead of the center is because the difference between short and long words were often large, so fixations were often mapped inaccurately when using the centers of short words.

4.2 Feature extraction

This section describes features that we use in the proposed method.

first gaze duration on word w_x:
: is the first gaze duration mapped on word w_x. If w_x is skipped once, we considered the regressive gaze duration as the first gaze duration.

number of fixations $n(w_x)$:
: total number of fixations occurred on word w_x.

number of regressions on word w_x:
: total number of regressive fixations occurred on word w_x.

$$reg(w_x) = \left|\{\, k \in [2, m] \,|\, w(k-1) > x, w(k) = x \,\}\right|. \quad (3)$$

$w(k)$ is a function to determine on which word the kth fixation is mapped. When $w(k) = x$, kth fixation is mapped on the xth word.

mean fixation duration on word w_x:
: is averaged duration of all fixations which are assigned to word w_x.

maximum gaze duration on word w_x:

$$g_{\max}(w_x) = \max\{G(w_x)\}. \quad (4)$$

A sequence of gazes on word w_x is given as $G(w_x) = \{g_1(w_x), g_2(w_x), ..., g_L(w_x)\}$.

pupil diameter variation on word w_x:

$$p_{\text{variation}}(w_x) = p_{\max}(w_x) - p_{\min}(w_x). \quad (5)$$

$p_{\max}(w_x)$ is the maximum pupil diameter on word w_x and $p_{\min}(w_x)$ is the minimum pupil diameter.

word length: $len(w_x)$

Pixel length of word w_x on a 1600x900 screen:

$$len(w_x) = D(s_x, e_x). \quad (6)$$

where $D(s_x, e_k)$ is a function to calculate Euclid distance between the xth word w_x's start point s_x and end point e_x.

4.3 Feature selection

We first attempted to use all features for detection but some of the features were not as effective as we predicted. To find effective features to detect unknown words accurately, we calculated correlation coefficients and p-values according to student's t-test for each of the eye movement and linguistic features. The t-test measured whether the mean values of unknown and known words were significantly different. According to the correlation coefficients and t-test, we selected effective features. We also normalized every feature to reduce the difference of units in each feature.

4.4 Classifier

We used SVMs with RBF kernels as classifiers. An evaluation was performed by 10-fold cross validation (for personalized models) and leave-one-subject-out cross validation (for non-personalized models). As our evaluation measure, we used F-measure because our recorded data was biased (unknown samples: 640 words and known samples: 9350 words). We used grid-search to find optimized parameters for each model.

5. EXPERIMENTAL EVALUATION

5.1 Eye movement recording

We recorded eye movement data while subjects read a stimulus document on a screen. While reading on the screen, the head-movements of participants were not conspicuous. Therefore, we used a table-mounted eye tracker because it can obtain focal points with high resolution accuracy when head movement is relatively small [2].

5.2 Procedure

We recruited a total of 12 Japanese graduate students (9 males and 3 females) who can read English (assessed by the TOEIC [15] English skill assessment score.). Each participant gave informed consent before the data recording. Demographics of all participants are shown in Table 1. The participants first entered an experiment room, and read three types of English documents while the eye-tracker recorded eye movement. After recording, participants were directed to manually annotate unknown words. Participants were asked to summarize the documents to ensure that they made an attempt to read and understand the material. As the visual stimuli, we used documents from Japanese entrance English examination book entitled "sokudokueitango" by H.Kazahaya, TOEFL[2] sample examination, and an academic paper by Rotter [11]. The visual stimulus documents were displayed on a screen with 6 line breaks to avoid vertical confounds (Figure 2). Every document was displayed as a PDF file. There were less than 4 sentences in each page. As the eye-tracker, we used a table-mounted, Tobii pro x2-30. To calculate word rarity, We used TreeTagger [12] to extract stems of verbs and nouns from each word. For those stems, we extracted word frequency ranking from the BNC word frequency lists [6], then we determined $rare(w_x)$ based on equation (3).

Table 1: Participants information. P: Participants' ID, UW: Number of unknown words, UW rate: unknown word rate in experimental document, ET: time elapsed to read.

	UW	UW Rate	ET [s]	TOEIC
P1	72	0.076	1909.7	475
P2	122	0.129	2418.9	530
P3	72	0.076	656.6	560
P4	91	0.096	1309.0	675
P5	60	0.063	564.7	680
P6	75	0.079	551.8	685
P7	42	0.044	865.1	740
P8	36	0.037	918.0	745
P9	39	0.041	765.0	785
P10	31	0.033	460.8	790
Total	640	0.067	10419.5	

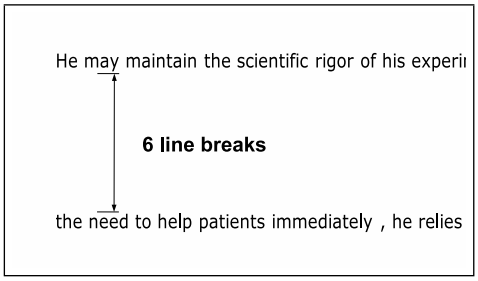

Figure 2: An example of the stimulus document.

Each participant read a total of 949 words and annotated whether each word was unknown or known on printed paper. We removed 2 participants who had fatal errors of the eye tracker during recording and finally 640 out of 9490 words were annotated as unknown.

5.3 Results

5.3.1 Feature selection

Table 2 shows the ranking of the features according with correlation with the known/unknown tags. The results indicate that all null hypotheses are rejected ($p < 0.01$). Therefore, all features were related to word labels. Max gaze duration was the most correlated eye gaze feature as measured by correlation coefficients and p-values.

5.3.2 Classification

[2]http://www.ets.org/toefl/

Table 2: Feature selection ranking by correlation coefficient r.

Feature	r	p-value
word rarity : $rare(w)$	0.462	e-233
word length : $len(w)$	0.339	e-171
max gaze duration : $g_{\max}(w)$	0.318	e-53
first gaze duration :$g_{\text{first}}(w)$	0.265	e-15
total gaze duration :$g_{\text{total}}(w)$	0.252	e-45
number of fixation :$n(w)$	0.208	e-40
mean fixation : $f_{mean}(w)$	0.153	e-31
pupil variation: $p_{\text{variation}}(w)$	0.136	e-19
regression : $reg(w)$	0.109	e-16

Table 3 shows a comparison of classifiers and eye gaze features according to 10-fold cross validation. As a baseline, we considered the threshold function as mentioned in equation (4). The results showed that the proposed method using SVM with max gaze duration and word length features improved performance as measured by F-measure. We applied the paired-bootstrap [7] test to confirm statistical significance, and we found that the improvement between the baseline and proposed learning method using SVMs was a significant ($p < 0.01$). In addition, use of max gaze duration provided significant improvement (paired boot strap test between SVM with total gaze and SVM with max gaze: $p < 0.01$). We also confirmed the SVMs with max gaze duration, word rarity, and word length had the best performance (F-measure: 0.556). We confirmed that results remain at 0.556 by randomizing the folds several times (0.556 $\pm$ 0.001).

Table 3: Classification results.

Features	$rare$ g_{total}	$rare$ g_{total}	$rare$ len	$rare$ $g_{\max}$	$rare$ $g_{\max}$ len
classifier	Baseline	SVM	SVM	SVM	SVM
Precision	0.482	0.457	0.501	0.472	0.507
Recall	0.375	0.630	0.581	0.639	0.616
False Positive	0.027	0.054	0.042	0.052	0.043
F-measure	0.422	0.530	0.538	0.543	0.556

5.3.3 Subject dependency

Next, we examined the effect of subject dependency using the best model with max gaze duration, word rarity, and word length features had a best performance. The result of leave-one-subject-out cross validation was compared to the result of 10-fold cross validation described in previous section. Table 4 shows that with subject dependency, the detection performance becomes slightly better (0.550 to 0.556). This indicates personalization is effective to improve the detection performance.

6. DISCUSSION

We proposed several eye gaze features and classifiers to detect unknown words. Using the most effective eye gaze features and SVMs with RBF kernels, we achieved a significant improvement in detection as measured by F-measure from the 0.422 of a baseline to 0.556. SVMs with max gaze duration, word rarity, and word length provided the highest performance. In addtion, max gaze duration can significantly contribute to improve detection performance.

Table 4: The effect of subject dependency for SVM with features: $rare/g_{\max}/len$.

Dependency	No	Yes
Precision	0.502	0.507
Recall	0.608	0.616
False Positive	0.038	0.043
F-measure	0.550	0.556

Moreover, we found personalization yet improves the detection performance. One limitation of our study is that, it has shown that pupil size diameter change needs 200-300ms latency to reflect human perception [4]. However, this latency was not considered in our approach. Also, fixation-saccade classification was performed by default settings of the Tobii studio IV-filter. Hence, we would like to consider other appropriate fixation-saccade classification algorithm in future work.

7. ACKNOWLEDGMENTS

This work was supported by KAKEN 26540117 and 16K16172.

8. REFERENCES

[1] C. Cortes and V. Vapnik. Support-vector networks. Machine learning, 20(3):273–297, 1995.

[2] A. Duchowski. Eye tracking methodology: Theory and practice, volume 373. Springer Science & Business Media, 2007.

[3] Y. Ehara, N. Shimizu, T. Ninomiya, and H. Nakagawa. Personalized reading support for second-language web documents by collective intelligence. In Proceedings of the 15th international conference on Intelligent user interfaces, pages 51–60. ACM, 2010.

[4] B. Gagl, S. Hawelka, and F. Hutzler. Systematic influence of gaze position on pupil size measurement: analysis and correction. Behavior research methods, 43(4):1171–1181, 2011.

[5] A. Hyrskykari. Eyes in attentive interfaces: Experiences from creating iDict, a gaze-aware reading aid. Tampereen yliopisto, 2006.

[6] A. Kilgarriff. BNC database and word frequency lists. HTTP:< http://www. kilgarriff. co. uk/bnc-readme. html>(accessed 24 Feb 2016), 1995.

[7] P. Koehn. Statistical significance tests for machine translation evaluation. In Empirical Methods in Natural Language Processing (EMNLP), pages 388–395. Citeseer, 2004.

[8] K. Kunze, H. Kawaichi, K. Yoshimura, and K. Kise. Towards inferring language expertise using eye tracking. In Computer-Human Inter- action (CHI)'13 Extended Abstracts on Human Factors in Computing Systems, pages 217–222. ACM, 2013.

[9] P. Martínez-Gómez and A. Aizawa. Recognition of understanding level and language skill using measurements of reading behavior. In Proceedings of

the 19th international conference on Intelligent User Interfaces, pages 95–104. ACM, 2014.

[10] K. Rayner. Eye movements in reading and information processing: 20 years of research. Psychological bulletin, 124(3):372, 1998.

[11] J. B. Rotter. Social learning and clinical psychology. 1954.

[12] H. Schmid. Probabilistic part-of-speech tagging using decision trees. In Proceedings of the international conference on new methods in language processing, volume 12, pages 44–49. Citeseer, 1994.

[13] L. E. Sibert and R. J. Jacob. Evaluation of eye gaze interaction. In Proceedings of the SIGCHI conference on Human Factors in Computing Systems, pages 281–288. ACM, 2000.

[14] C. Ware and H. H. Mikaelian. An evaluation of an eye tracker as a device for computer input2. In ACM SIGCHI Bulletin, volume 17, pages 183–188. ACM, 1987.

[15] P. E. Woodford. The test of english for international communication (TOEIC). 1980.