Quantified Reading and Learning for Sharing Experiences

Koichi Kise
Osaka Prefecture University
kise@cs.osakafu-u.ac.jp

Olivier Augereau
Osaka Prefecture University
augereau.o@gmail.com

Yuzuko Utsumi
Osaka Prefecture University
yuzuko@cs.osakafu-u.ac.jp

Masakazu Iwamura
Osaka Prefecture University
masa@cs.osakafu-u.ac.jp

Kai Kunze
Keio University
kai.kunze@gmail.com

Shoya Ishimaru
German Research Center for Artificial Intelligence (DFKI)
shoya.ishimaru@dfki.de

Andreas Dengel
German Research Center for Artificial Intelligence (DFKI)
andreas.dengel@dfki.de

UbiComp/ISWC'17 Adjunct , September 11–15, 2017, Maui, HI, USA
ACM 978-1-4503-5190-4/17/09.
https://doi.org/10.1145/3123024.3129274

Abstract
This paper presents two topics. The first is an overview of our recently started project called "experiential supplement", which is to transfer human experiences by recording and processing them to be acceptable by others. The second is sensing technologies for producing experiential supplements in the context of learning. Because a basic activity of learning is reading, we also deal with sensing of reading. Methods for quantifying the reading in terms of the number of read words, the period of reading, type of read documents, identifying read words are shown with experimental results. As for learning, we propose methods for estimating the English ability, confidence in answers to English questions, and estimating unknown words. The above are sensed by various sensors including eye trackers, EOG, EEG, and the first person vision.

Author Keywords
e-learing; eye-tracking; EOG; confidence; TOEIC; unknown words; wordometer; reading detection; read word identification; human experience

ACM Classification Keywords
H.5.m [Information interfaces and presentation (e.g., HCI)]: Miscellaneous; K.3.m [Computers and education]: Miscellaneous

Introduction

Human experience is valuable information that can be produced only by humans. As the proverb “Experience is the best teacher” tells us, humans can extend their ability by gaining experiences. We have just launched a project called “experiential supplement” to help humans gain experiences by “supplements” generated from other human experiences. In other words, experiences are shared through supplements.

In this project, we have three main application fields: learning, health care, and sports/entertainment. In this paper, we focus on learning and report what we have achieved about learning and reading which is a basic activity of learning.

In order to share experiences, it is necessary to capture and describe them. Since experiences are produced by humans in various ways, they should be sensed in many aspects. In this paper, such digitized activities of reading and learning are called “quantified reading and learning.” The question here is what and how to sense. This paper summarizes our trials of sensing from quantity to quality of reading and learning. To be precise, we describe (1) quantifying the amount of reading in terms of the number of words, (2) detection of reading, (3) classifying reading in terms of document types, (4) the precise description of reading by identifying read words, (5) estimation of English ability through analysis of behavior answering to English questions, (6) estimation of confidence of answers to English questions, (7) estimation of unknown words by analyzing reading activities.

In the following we first describe the overall project in the next section before introducing the above 7 methods. The final section summaries the paper.

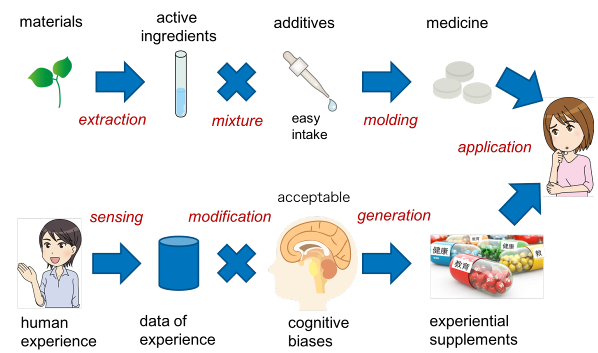


Figure 1: Medicine and experiental supplements.

Experiential Supplements

The notion of experiential supplements can be described by analogy with production of medicine or chemical supplements as shown the upper part of Fig. 1. Materials are processed to extract active ingredients, which are then mixed with additives for easy intake. The next step is to mold medicine which will be applied to a person. The case of experiential supplements is shown in the lower part. Human experiences are sensed to acquire data of experiences. Then the data are modified with additives to make it acceptable by other people. As additives for experiences, we employ various cognitive nature such as cognitive biases which can modify the attitude of the user. This allows us to generate a modified information about experiences called “experiential supplements” and to prescribe them to the user.

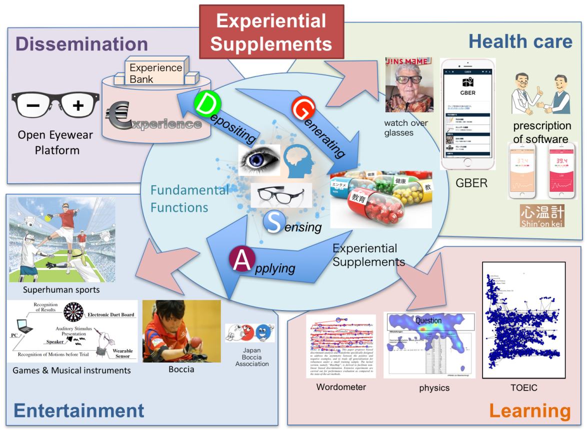


Figure 2: Overview of experiential supplements.

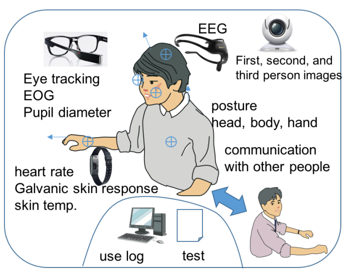


Figure 3: Deep sensing of reading and learning.

Figure 2 shows the overview of the project whose purpose is to develop a platform for changing human behaviors in a better way by applying human experiences. The underlying notion is that in many cases a problem a user is facing has already been experienced and solved by others. The overall processing is as follows. First, human experiences are sensed to produce the data of experiences, which are stored in the database called the experience bank. Upon request from the user the data of experiences are selected and given to the process of generating experiential supplements. In the process of generation, we employ cognitive biases to convince the user to take into account the generated supplements. For example, by using a social bias the user could be more sensitive to his/her position in his/her group. This motivates the user to improve his/her performance. A memory bias could help the user memorize the item to learn more efficiently. Then they are applied to the user to change his/her behavior. In order to achieve a better application, we need to consider when and where the experiential supplement should be given. The notions of interruptability and acceptability of information play an important role in this context. We are planning to work on these important research issues in a few years, after finishing the research on the sensing part.

As the application fields, we have the following three: learning, health care and sports/entertainment. As compared to learning, which is mainly cognitive activities, sports / entertainment is mostly physical activities, and health care is the mixture of both depending on the type of health we focus.

In the following section we focus on the application field of learning and describe the details of its sensing part.

In order to realize personalized, tailor-made learning, various e-learning systems have already been proposed and used. In many cases, their sensing part is simple: they employ system usage logs and results of tests to know the users. However, they are not enough to know details of internal states of the users such as emotion, confidence and concentration. We propose to use a wider variety of sensors as shown in Fig. 3. As compared to the conventional way of sensing, which we call shallow sensing, we call sensing using those various sensors “deep sensing.”

Quantified Reading and Learning

In this section, we describe the details of sensing in the field of learning and some examples of tasks and methods we have already developed. The tasks we consider here are to assist English learning from low to high levels as shown in Fig. 4. Since all of the tasks are about reading, which have been well studied in relation to eye movements [9, 10], we employ eye trackers as a main sensor.

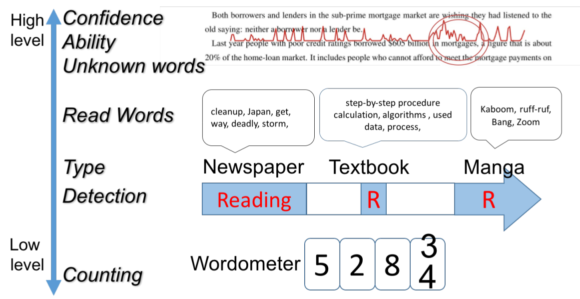


Figure 4: Levels of tasks.

Wordometer

Let us start with a low level sensing of quantity. The most fundamental function is “wordometer” which measures how many words the user has read [6]. We intend to build it as a step counter (pedometer) of our knowledge life.

The wordometer has been implemented using one of the three devices: JINS MEME, mobile and stationary eye trackers. JINS MEME is commercial glasses with EOG (electrooculography), accelerometer and gyroscope[1]. With this device, we are able to detect eye movement, blinks and head motion. As eye trackers, we employ a mobile eye tracker (SMI Eye Tracking Glass (ETG) [2]) and a stationary eye tracker (Tobii Eye X [3]).

We have several versions of wordometers depending on the device and the method to use. However, the basic computation is shared as follows. The task of estimating the number of read words is formalized as a regression problem. Taking features of sensor outputs as input, we apply support vector regression to estimate the number of words. First, raw eye gaze data are transformed into a sequence of saccades and fixations [3]. Then some features such as the number of detected line breaks, and the number of fixations are calculated. The line break is detected as a large backward (right to left) movement of eyes.

The error of estimation depends on the device to use, setup of the experiments and the amount of text to read. The setup indicates whether the learning by support vector regression is either user dependent or independent, as well as document dependent or independent. Generally speaking, the error ranges from 3% to 14 % [5, 6, 4], which is comparable to that of pedometers.

We expect that, as pedometers motivate people to walk more and continue to walk, the same can happen to users of wordometers. We are now trying to prove it experimentally with a larger scale user study.

Reading Detection

Another basic function in terms of quantity is reading detection to know the period of time the user read. The reading behavior is characterized by periodical eye movement: a sequence of small forward saccades (from left to right) followed by a large backward saccade (line break).

As a device, we have employed one of the following: JINS MEME and SMI ETG. The method of reading detection is quite different depending on the device to use.

Let us explain here the case with SMI ETG. We employ eye movement features such as the average and the variance of saccade length, the sum and the average of fixation time, etc. After user independent learning of SVM to classify ac-

[1] https://jins-meme.com/en/academic/
[2] https://www.smivision.com/
[3] http://tobiigaming.com/product/tobii-eyex/

Figure 6: Read word identification and its representation as tag clouds. Two persons read the same document but the results are quite different.

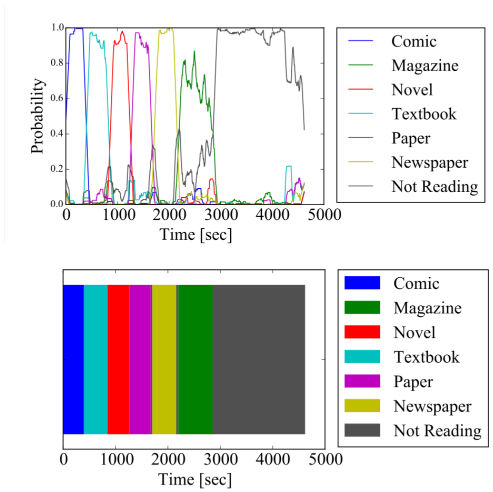


Figure 5: Detection of reading and document type recognition.

tivities into either reading or others, the F-measure 90% has been obtained from the experiment with 10 people doing 50 to 80 minutes activities half of which are reading.

Document Type Recognition

In order to know more about the period of reading, we apply document type recognition. This allows us to obtain statistics of reading time for each document type. As types to be recognized, we employ the following six: newspaper, textbook, scientific paper, novel, comic and magazine. We have employed one of the following: an electroencephalograph (EEG) device called Emotiv [7], and SMI ETG [8, 12].

Let us describe a method with SMI ETG to make the document type recognition user independent [12]. The each frame image of the video captured by the scene camera on ETG is segmented based on the eye gaze and passed to the classifier using GoogLeNet [13]. This allows us to produce the probability about which type of document the user reads at each frame image as shown in the upper part of Fig. 5. This classification includes the class of "not reading" so that reading detection is also done.

This method was evaluated through the experiment with 8 subjects and 15 minutes read for each class. As a result of reading detection, we obtained recall 92% and precision 95%, which are both high enough. As for the document type recognition, recall 82% with precision 85% have been achieved. The typical errors are caused by similar images from papers and text books, etc. To distinguish such cases, it is necessary to include analysis of contents.

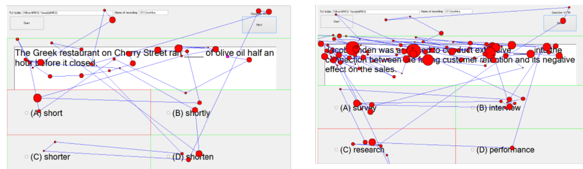
Read Word Identification

We can consider further quantification of reading activities by identifying read words. This allows us to build a Bag-of-Words (BOW) model of our reading activities from which we can know the contents of learning.

We employ one of the two eye tracking devices Tobii Eye X and SMI ETG for this purpose. Let us focus here on Tobii Eye X. The main issue is how to deal with errors on the estimated fixation positions [11]. After the matching of eye gaze data to the displayed textlines, it is possible to know the start and the end of reading. This enables us to build the BOW illustrated as tag clouds shown in Fig. 6 [2]. The weight for each word to draw the tag cloud is calculated by taking into account the location and duration of fixations.

Confidence

It is helpful for learners and teachers if the confidence of answers is estimated to determine which part should be reviewed. In particular, it is meaningful to find correct answers without confidence and incorrect ones with high confidence. This is because the former may be by chance, and the latter



(a) with confidence (b) without confidence

Figure 7: Estimation of confidence on multiple choice questions.

indicates serious misunderstanding.

In general, it is not easy to solve this problem. However, it is tractable if we limit the estimation to answers of multiple choice questions. This is because, as shown in Fig. 7, eye gaze for an answer with confidence is typically different from that without it.

We have developed a method to estimate whether the answer is with confidence or not by using Tobii Eye X. As the domain for application, we employ an English standardized test called TOEIC (Test of English for International Communication) whose score ranges from 10 to 990. In the TOEIC test, we focus on Part 5 of reading section: a sentence completion task with 4 choices to fill the blank. The SVM is employed for this purpose with gaze features as input in addition to the correctness of the answer. We employed as subjects 11 Japanese university and graduate school students whose TOEIC score ranged from 450 and 940. Based on the experiment with 80 questions, it has shown that our method is capable of estimating the confidence with the accuracy of 90.1% as shown in Fig. 8.

English Ability

The wordometer is to count the quantity of reading to encourage learners to read more. We consider that the same

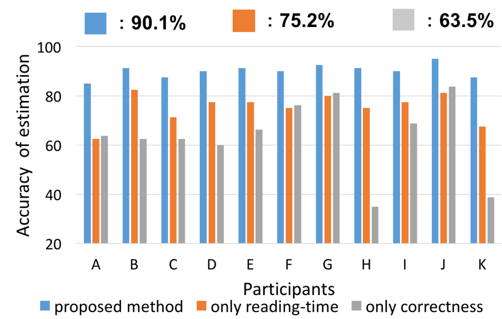


Figure 8: Results of estimation of confidence. Gray bars correspond to the estimation by assuming that all correct answers were with confidence. Orange ones indicate the estimation based upon the time needed for answer; the longer the less confident. As compared to them, the proposed estimation represented as blue bars is much superior.

holds for the English ability if we can quantify it. For this purpose, we estimate the TOEIC score [1]. In this task, we focus on the Part 7 (passage comprehension) of the reading section. A typical format of the Part 7 includes a passage and some questions about its contents, each of which is a four-choice question.

We analyzed the eye gaze data on 15 documents read by 16 subjects who are Japanese, university and graduate school students whose TOEIC score ranged from 390 to 945. As the device, we employed the SMI stationary eye tracker RED250. Figure 9 shows an example eye gaze.

The accuracy of estimation depends on the setting of experiments: whether learning is user independent or not, as well as document independent or not. In the easiest case, that is document dependent setting with optimal feature

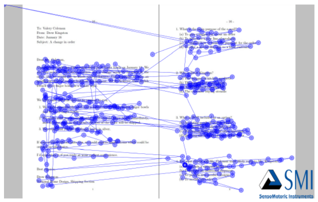


Figure 9: Estimation of TOEIC scores.

selection, we can estimate the score with the error of 30 points by taking into account the eye gaze for two documents as shown in Fig. 10.

Unknown Words
Vocabulary building is a basic task for English learners. However, it requires a labor-intensive work of making a list of unknown words. It also disturbs the rhythm of reading and understanding. In order to cope with this problem, we have built a method that automates the identification of unknown words by taking as input only the gaze data while reading English text. As the device, we used theTobii Eye X. A deep neural network is employed to determine whether a word with the gaze is unknown. Figure 11 shows experimental results with 5 subjects (Japanese, university and graduate school students) and 16 documents. As compared to estimation using word frequency (less frequent words are supposed to be difficult), the proposed method is able to produce better results.

Conclusion

In this paper, we summarized our results of quantify reading and learning behaviors using various devices including eye trackers, EOG, and the first person vision. The purpose of sensing is to know the internal states of readers and learners from low to high levels.

As we mentioned in Section 2, sensing is just the first step towards sharing learning experiences. It is required to build methods to describe and share the experiences in the form called “supplements” to change learners’ behavior. We plan to complete this project within five years and hope that we will have a chance to explain the latter steps.

Acknowledgment
We would like to thank Charles Lima Sanches, Kensuke Hoshika, Yuki Shiga, Hiroki Fujiyoshi, Ayano Okoso for

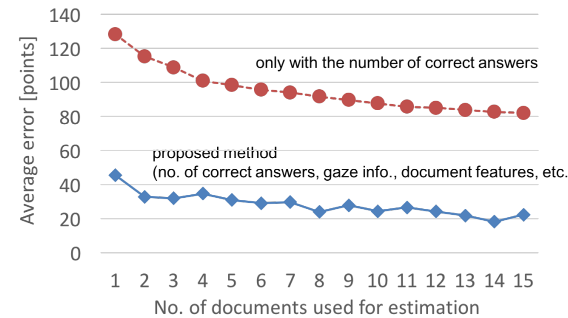


Figure 10: Results of TOEIC score estimation. As compared to the estimation by only using the number of correct answers, the proposed method offers estimation with much less errors.

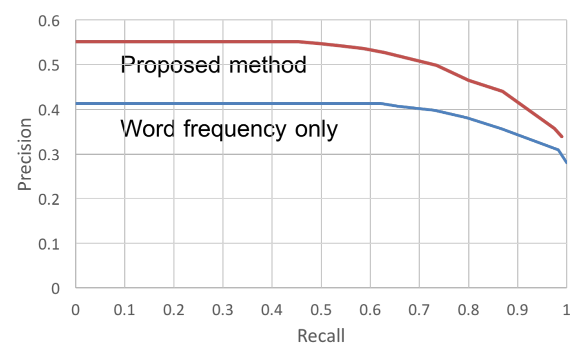


Figure 11: Results of the estimation of unknown words.

their work on their research topics included in this paper. This research was in part supported by JST CREST (JP-MJCR16E1) , JSPS Grant-in-Aid for Scientific Research (15K12172), and Key Project in Osaka Prefecture Univ.

REFERENCES

1. Olivier Augereau, Hiroki Fujiyoshi, and Koichi Kise. 2016. Towards an Automated Estimation of English Skill via TOEIC Score Based on Reading Analysis. In *Proc. International Conference on Pattern Recognition*.
2. Olivier Augereau, Koichi Kise, and Kensuke Hoshika. 2015. A Proposal of a Document Image Reading-life Log Based on Document Image Retrieval and Eyetracking. In *Proc. International Conf. on Document Analysis and Recognition (ICDAR2015)*. 23–26.
3. Georg Buscher and Andreas Dengel. 2009. Gaze-based filtering of relevant document segments. In *Proc. International World Wide Web Conference (WWW)*. 20–24.
4. Shoya Ishimaru, Kai Kunze, Koichi Kise, and Andreas Dengel. 2016. The Wordometer 2.0 - Estimating the Number of Words You Read in Real Life using Commercial EOG Glasses. In *Proc. UbiComp2016 Adjunct*. 1217–1220.
5. Kai Kunze, Hitoshi Kawaichi, Koichi Kise, and Kazuyo Yoshimura. 2013. The Wordometer - Estimating the Number of Words Read Using Document Image Retrieval and Mobile Eye Tracking. In *Proc. ICDAR2013*. 25 – 29.
6. Kai Kunze, Katsutoshi Masai, Masahiko Inami, Ömer Sacakli, Marcus Liwicki, Andreas Dengel, Shoya Ishimaru, and Koichi Kise. 2015. Quantifying Reading Habits – Counting How Many Words You Read. In *Proc. UbiComp2015*. 87–96.
7. Kai Kunze, Yuki Shiga, Shoya Ishimaru, and Koichi Kise. 2013a. Reading Activity Recognition Using an Off-the-Shelf EEG — Detecting Reading Activities and Distinguishing Genres of Documents. In *Proc. International Conf. on Document Analysis and Recognition (ICDAR2013)*. 96–100.
8. Kai Kunze, Yuzuko Utsumi, Yuki Shiga, Koichi Kise, and Andreas Bulling. 2013b. I know what you are reading: recognition of document types using mobile eye tracking. In *Proc. International Symposium on Wearable Comuters*. 113–116.
9. Keith Rayner. 1998. Eye movements in reading and information processing: 20 years of research. *Psychological bulleti* 124, 3 (1998), 372–422.
10. Keith Rayner, Alexander Pollatsek, Jane Ashby, and Calrles Clifton Jr. 2012. *Psychology of Reading* (2nd ed.). Psychology Press.
11. Charles Lima Sanches, Olivier Augereau, and Koichi Kise. 2016. Vertical Error Correction of Eye Trackers in Nonrestrictive Reading Condition. *IPSJ Trans. on Computer Vision and Applications* 8, 7 (2016).
12. Yuki Shiga, Yuzuko Utsumi, Masakazu Iwamura, Kai Kunze, and Koichi Kise. 2016. Automatic Document Type Classification Using Eye Movements and Egocentric Images. *Trans. IEICE (D)* J99-D, 9 (2016), 950–958.
13. Christian Szegedy, Wei Liu, Yangqing Jia, Pierre Sermanet, Scott Reed, Dragomir Anguelov, Dumitru Erhan, Vincent Vanhoucke, and Andrew Rabinovich. 2015. Going Deeper with Convolutions. In *Proc. CVPR 2015*. IEEE.