ISASE 2019

Proposal for CAPTCHA that Makes Use of the Human Ability to Understand Context

Jue ZHANG*, Masahiro MORITA*, Brian Henson** and Bryan Matthews***

** Kogakuin University, 1-24-2 Nishi-Shinjuku, Shinjuku-ku, Tokyo 163-8677, Japan*
zhangjue@cc.kogakuin.ac.jp
*** School of Mechanical Engineering, University of Leeds, Woodhouse Ln, Leeds LS2 9JT UK*
B.Henson@leeds.ac.uk
**** Institute for Transport Studies, Geography East Building, University of Leeds, LS2 9JT UK*
b.matthews@its.leeds.ac.uk

Abstract: For the visually impaired, it is difficult to use CAPTCHA that was generated using visual information. The only way is to install a browser add-on and, by eliminating CAPTCHA, to collect information from the internet. However, it is not easy for visually impaired people to install an add-on, and there will also be disadvantages for general users if anyone is able to eliminate the CAPTCHA function. As such, using text that can be converted into voice, we proposed a sentence rearranging format of CAPTCHA that makes use of the human ability to understand context. Through experiments, we examined the kind of sentences, number of Japanese characters, number of sentences, and special characteristics of sentences that the sentence rearranging format could accommodate. Our conclusion was that, primarily, humans find it easy to understand context to the extent that the subject of the sentence is clearly indicated and is linked to the content of the sentence, that it is appropriate to use four sentences or fewer as multiple-choice options in problem sentences, and that less than 100 Japanese characters is ideal for the number of characters.
Keywords: *CAPTCHA, Visually impaired, Context understanding*

1. INTRODUCTION

As the internet has developed in recent years, DoS (Denial of Service) attacks, using automated programs to conduct large-scale unauthorized usage requests or large scale unauthorized account acquisitions, have become a problem for Web services such as free Web mail and blogs. In order to deal with this problem, the importance of the testing method called CAPTCHA (Completely Automated Public Turing test to tell Computers and Humans Apart) as a countermeasure is increasing. At present, the main methods used for CAPTCHA are either a method of displaying an image with letters that have been distorted or to which noise has been added, or else a method of choosing an image that agrees with displayed requirements. These methods rely on only visual information, and it is extremely difficult for the visually impaired to answer. As a substitute for methods that rely on only visual information, CAPTCHA that uses a transformed voice has also been implemented. However, it has been pointed out that it is difficult not only for machines but also for humans to discern the correct answer [1][2][3][4].

To solve this problem, in this research we proposed a sentence rearranging format of CAPTCHA that focused on the human characteristic of being able to understand the flow of sentences, and which used text that could be used by both the visual and the auditory senses. Moreover, as a first step, we also analyzed and examined the requirements for suitable sentences.

2. PROPOSED METHOD

2.1 Examples from previous research

In previous research, a method using sentences called "word salads" was proposed as a method for CAPTCHA that used text [4]. Word salads are machine generated sentences with randomly combined words and conjunctions that are correct grammatically but that have unnatural content. In the corresponding method, machines and humans are distinguished by having the user discriminate between sentences. As a requirement for a problem sentence to function as CAPTCHA, it is desirable that fresh problem sentences are continually increasing in number. In the algorithm for generating problem sentences in the word salad method, sentences newly generated everyday on SNS are collected by a program, and fresh problems can be generated using a

Marcov generation model to generate natural sentences and unnatural sentences.

In addition, a method for using the discomfort felt from machine translations was proposed as another method for CAPTCHA that used text [5]. In this method, an image is first displayed and a person is made to input a sentence explaining this image. If the input sentence can be confirmed as natural, then this sentence is stored in a database. To generate sentences for the test, machine translation is repeatedly applied to the original sentences stored in the above database, thereby creating sentences that are grammatically correct but that only humans would feel to be somewhat wrong. Humans and computers would be distinguished by having the user choose the natural sentence from a number of multiple-choice options in which the natural sentence was mixed with unnatural sentences. Because fresh sentences are generated by using a database, the requirement of producing new problems, and the requirement of the problem sentences having the property of distinguishing humans from computers, are both fulfilled.

On the basis of the methods proposed in the previous research described above, in this study we propose a new processing algorithm that uses text in the same way as before to produce an easier CAPTCHA method.

As a common point among CAPTCHA methods that use text, the sentences are grammatically correct but contain elements that seem unnatural only to humans, due to things like meaning, context, or structure. In this way, they distinguish between humans and computers by making use of the human ability to understand. In this research, we place the emphasis on the ability to understand context, and we propose a method for rearranging sentences.

2.2 Requirements for problem sentences

When generating problems for CAPTCHA that uses text, there is the requirement of automatically acquiring original text, and there is also the requirement of being able to automatically generate problems from the original text so acquired.

For the original text used for problem sentences in this study, we used a digital library for material, in which the passages used were assumed to have continuous content. Automatically acquired original text is turned into sentence rearranging problems with single sentences as the unit.

In addition, there are various requirements when generating problem sentences.

First, it is desirable that there should be only one unique contextual order. This is because it might be difficult even for humans to find the correct answer if a meaning can be understood even after the order of the sentences has been rearranged.

Second, it is desirable that the amount of information be within the range that visually impaired people will be able to remember if the sentences are read by a voice. This is because, compared to a healthy user who can always identify the problem sentences being displayed, for visually impaired people who can only receive the information of the sentences by voice, and who would also find it difficult to take notes, it will be necessary for them to solve the problems from within the range of their memory.

As factors that influence the above requirements, there is first the amount of information in the passage, and second, there is the content of the passage. As factors that constitute the content of the passages for contextual understanding, on the side of grammar there are things like subjects and verbs, and on the side of knowledge there is the difficulty of words and expressions.

2.3 The sentence extraction algorithm

To generate problem sentences for this experiment, the sentences used were assumed to have continuous content. In order to study the requirements of passages that are suitable for this CAPTCHA method, in addition to the above mentioned influencing factors, it is necessary to use sentences that have distinguishing features or multiple ways of being written. For this reason, in this research we used a sample of novels from the Aozora Bunko, from the following area: "9 Literature > 91 Japanese Literature > 913 Novels/Stories," to find material that fulfils these requirements.

In the selection of novels, we established three conditions for considering the special characteristics of sentences suitable for CAPTCHA: (1) number of sentences; (2) number of Japanese characters; and (3) genre of novel. We determined passages by using random numbers in the order of author number and number of works on that same website, until passages that fulfilled each condition came out. Moreover, in the creation of problem sentences, we decided that, in principle, one sentence would be counted from the previous sentence until the appearance of a ".", "new line," "?", or "!" However, we decided that ".", "?", and "!" that appeared within quotations or brackets would not be viewed as the end point of a sentence. This was because we judged that in sentences with the above conditions, these punctuation marks were not ending one sentence but were rather

indicating the content of someone's speech, and it was possible that multiple sentences could be generated within a single set of quotation marks, or else that there would be many patterns in which immediately after the sentence in quotations, "he said" would be attached as a short sentence. If each of these were counted as one sentence and used to generate problem sentences, then it would be possible that the comprehension of the meaning of the sentences would be obstructed, insofar as this could easily give rise to patterns in which there might be multiple valid interpretations of the sentence order.

After determining the novels from which to extract, as a process common for all conditions, the question of whether the beginning of the extracted passage should be the nth sentence was determined according to the range of ($1 \leqq n \leqq 30$), and from the nth sentence three to five sentences that fulfilled the conditions were extracted and the order of the sentences was then randomly rearranged. Moreover, with regard to the range of n, it was desirable that ($1 \leqq n \leqq$ the number of sentences in the extraction interval -(the number of sentences extracted-1)). However, because this time the extraction was done manually, we designated this range in order to have a range in which it would be easy to conduct a precise count and in order that this compromise would overlap with the random extraction of sentences.

As a procedure for selecting unique novels and problem sentences that fulfilled the conditions, (1) concerning the number of sentences, we selected two novels divided into chapters, and we used random numbers to determine which sentences to extract from these chapters. We made the sample passages be from the same novel in order to makes sure the characteristics of the sentences matched, by making it the same novel among all the sample passages. In this way, we extracted from three chapters in each novel, for a total of six sample passages. After this, when the number of extracted sentences was 3, 4, or 5 sentences, we apportioned them using random numbers and followed the extraction procedures. (2) Concerning the number of Japanese characters, we first selected 20 novels using random numbers. After extracting three sentences, we calculated the total number of Japanese characters in each, according to three different patterns: up to 100 characters, up to 200 characters, and up to 300 characters. We then used random numbers to select one sentence from among those conforming to these three patterns. (3) Concerning the genre of the novels, we divided the novels that had not been used in (2), number of characters, into five different types depending on their content: "mystery/suspense," "fantasy/fairy tale," "youth/friendship," "history," "human drama." We then used random numbers to pick out one work from each of the genres, and we used three sentences as they were for problem sentences.

3. PASSAGE REARRANGING EXPERIMENT

3.1 Experimental method

With regard to the total of 14 written samples from which problem sentences were generated, we used Amazon Polly to change the written data into MP3 data. Then, after randomly rearranging the order for the test, we had 25 undergraduate/graduate students in their twenties, who had not read the novels from which the problem sentences had been extracted, listen either by speaker or by headphones to a voice reading the problem sentences, and then rearrange them back into the correct order. Moreover, even if the subjects were unable to answer after one attempt, in order to investigate the appropriate range for CAPTCHA concerning how many times it might take for the right answer to be given, we conducted the same experiment up to a maximum of three times for each experiment.

3.2 Experimental results and considerations

The average correct answer rate given by the experimental subjects up to their third response is shown in the table (table 1). Those with a correct response rate of less than 50% are shown in red, representing conditions that should be excluded in particular. Those with a correct response rate of more than 80% but less than 90% are shown in yellow, representing conditions that are somewhat appropriate. Those with a correct response rate of more than 90% are shown in green, representing conditions that are particularly suitable.

Table 1. The average correct answer rate up to the third response

(1) Number of extracted sentences					
3 sentencesA	4 sentencesA	5 sentencesA	3 sentencesB	3 sentencesB	3 sentencesB
0.36	0.84	0	0.96	0.8	0.2
(2) Number of Japanese characters					
up to 100 characters		up to 200 characters		up to 300 characters	
0.84		0.72		0.68	
(3) Genre of Novel					
Mystery Suspense	Fantasy Fairy tale	Youth Friendship	History	Human drama	
0.84	1	0.92	0.64	0.28	

Concerning condition (1), for both novel A and novel B, when the number of sentences goes from 4 to 5, the correct answer rate becomes significantly lower. The reason the correct answer rate in the 3 sentence problem from novel A is lower is due to the fact that there was weak contextual linkage between the multiple-choice

options of the problem sentences that were generated, such that even when the order of the sentences was rearranged, the experimental subjects could still understand a meaning. As such, there was a strong influence from factors beyond the number of sentences. From the above tendencies, it might be said that it would be appropriate to use 3 or 4 sentences as problem sentences.

Concerning condition (2), one can see that there is a tendency for the correct response rate to go down as the number of Japanese characters increases. From this result, it is desirable to choose sentences in which the number of characters is less than 100.

Concerning condition (3), "fantasy/fairy tale" and "youth/friendship" had a correct answer rate that surpassed 90%. In particular, all the experimental subjects were able to correctly answer problems in the "fantasy/fairy tale" category. In contrast, the correct answer rate for "human drama" was less than 50%, and "history" was second worst with another low rate. Concerning "fantasy/fairy tale," for which the correct answer rate was particularly high, we think that the flow of the sentences was easy to understand because the content was composed of words that even children would be able to understand, and because, if there was speech or actions, the subjects of the sentences were not omitted but were instead clearly indicated. With "youth/friendship," in the same way as with "fantasy/fairy tale," we think that the reason for the high success rate was that it is difficult to use hard words and expressions in this genre. Opposite to this, in "human drama," we think that the reason the correct answer rate was lower is that the structure of the passages was like a recitation, close to spoken language, with a low frequency of conjunction use and many sentence forms in which the subjects of the sentences were omitted. We think that the reason the correct answer rate was low in "history" was that many unfamiliar things appeared, such as the background of the historical epoch or historical words.

Considering the correct answer rate, the factor that had the strongest effect among the three conditions in general was the content of the sentences, even more than the number of characters of the number of sentences. In particular, there appears to be a tendency whereby the clearer the connection between the subject and its content for a given sentence, the easier the context becomes to understand.

In addition, the average number of responses under each condition for the factor that appears to have had the strongest influence on all respondents—condition (3)—is shown in table 2. The color scheme for each condition is still based on the applicability conditions for the CAPTCHA correct answer rates from table 1.

Except for "human drama," in which the correct answer rate was still low, the average number of times it took to give a correct answer in the other four categories was less than two, and in particular for "fantasy/fairy tale," for which everyone gave the correct answer, 99.6 of the experimental subjects gave the correct answer on their first try. Because it is desirable that, in CAPTCHA, the burden placed on experimental subjects be made as light as possible, we think that sentence formats close to the sentences of "fantasy/fairy tale" would be most suitable.

Table 2. Average number of responses for those who gave the correct response in requirement (3)

(3) Genre of Novel				
Mystery Suspense	Fantasy Fairy tale	Youth Friendship	History	Human drama
1.86	1.04	1.87	1.94	2.14

4. CONCLUSIONGS AND PROBLEMS FOR FUTURE RESEARCH

In this research, we proposed a method for rearranging sentences as a CAPTCHA method that does not rely only on visual information. Our goal was to clarify the conditions that would make sentences suitable for usage as sample sentences. We made problem sentences by using a sentence rearranging format, and we had experimental subjects provide answers. Our conclusion was that, primarily, humans find it easy to understand context to the extent that the subject of the sentence is clearly indicated and is linked to the content of the sentence, that it is appropriate to use four sentences or fewer as multiple-choice options in problem sentences, and that less than 100 Japanese characters is ideal for the number of characters.

The algorithm for the sentence rearranging formula proposed in this research was only for changing the order of the prepared linked sentences and returning them to the user. It can be expected that this program has the strong point of being easily adopted in a wide range of languages.

As problems for future consideration, from the perspective of refreshing problem sentences, it will be necessary to examine methods for original sentence collection in places other than digital libraries, and it will also be necessary to conduct an investigation on the influence of including noise in order to maintain the freshness of problems, which was not done in this experiment.

REFFERENCES

[1] Jeffrey P. Bigham and Anna C. Cavender. Evaluating Existing Audio CAPTCHAs and an Interface Optimized for Non-visual Use, CHI ’ 09, pp, 1829‐1838. ACM, 2009.

[2] Elie Bursztein, Steven Bethard, Celine Fabry, John C. Mitchell, and Dan Jurafsky, How Good are Humans at Solving CAPGCHAs’ a Large Scale Evaluation, SP ’ 10, pp. 399‐413. IEEE Computer Society, 2010

[3] Okamoto, Yamaguchi, Information security technology suitable for visually impaired medical system students, National University Corporation Tsukuba University of Technology, Vol. 21 (2) Mar. 2014

[4] Yamamoto, Okamoto, An Advanced CAPTCHA Using Generation of Accessible Verbal Questions, Computer Security Symposium 2013 pp21-23, 2013

[5] Yamamoto, J. D. Tygar, A Proposal of CAPTCHA using Strangeness in Machine Translation, Vol.2009-CSEC-46 No. 37, 2009