マルコフ連鎖による合成文章の不自然さを用いたCAPTCHAの安全性評価と改良について

鴨志田　芳典[†1]　　菊　池　浩　明[†1]

ボットによるアカウントの大量取得や，それに伴う不正行為への対策として広く用いられている CAPTCHA と呼ばれる機械判別方式は，コンピュータには判別が困難だが人間には容易である問題を利用する事でプログラムによる入力と人による入力とを識別する．我々は OCR 機能を持ったマルウェアやリレーアタックに耐性を持つ CAPTCHA として，ワードサラダと呼ばれるマルコフ連鎖による文章合成の不自然さを用いた CAPTCHA を提案している．本稿では提案手法への文章公正ツールを用いた攻撃に対する耐性の評価を行い，それに基づき提案手法の改良を行う．また提案手法が日本語以外の言語にも適用可能であるかを実験により評価し，その条件を考察する．

Evaluation of safety and improvement to the CAPTCHA using artificial synthesis sentences.

YOSHIFUMI KAMOSHIDA[†1]
and HIROAKI KIKUCHI[†1]

The CAPTCHA is a widely used technique to prevent malicious software, called bot, from obtaining false account. It uses an easy problem for human but difficult for machines to distinguish an input made be by a program and one by human. We have proposed a new CAPTCHA testing sentences synthesized from the Markov chain model, called "word salad", which is tolerant against malware with an OCR feature, or a "crowd sourcing" attack, called as "relay attack". In this paper, we estimate the degree of tolerance against the attack using proof-reading tool to distinguish the synthesized sentences, and present an improved protocol for the attack. Moreover, we show an experimental result of some languages other than Japanese.

1. 概　　要

近年，ボットによるアカウントの大量取得やそれに伴う不正行為への対策として CAPTCHA と呼ばれる不正プログラム判別方式が広く用いられている[1)]．CAPTCHA はコンピュータには判別が困難だが人間には容易である問題を利用する事で，ボットやエージェントなどのプログラムされた入力と人による入力とを識別する．しかしながら，最も広く用いられている文字列画像を変形させた CAPTCHA は，高精度の OCR 機能を持ったマルウェアによって破られてしまう事が報告されている[2)4)]．人間を使ったリレーアタックと呼ばれる攻撃も問題になっている[7)]．リレーアタックでは，攻撃者は CAPTCHA の問題を自分の運営する Web サイトに転載し，それを人に解かせることにより CAPTCHA を成功させる．転載された CAPTCHA を解く人間は発展途上国の低賃金労働者である．この様に，CAPTCHA はいずれ評価される可能性があるため，方式の多様性が求められている．

そこで，視覚的な情報だけではなく，人間のより高度な認知処理を用いた CAPTCHA の研究が行われている．Assira[3)] はコンピュータが画像の意味を理解する事の困難さを利用した代表的な CAPTCHA で，画面上に表示された複数の画像から，犬の画像か猫の画像のみを選択させる事により人間と不正プログラムを区別する．

我々の研究もこの試みのひとつであり，リレーアタックに耐性を持つ CAPTCHA として，ワードサラダと呼ばれるマルコフ連鎖よる文章合成の不自然さを用いた CAPTCHA を提案している[12)]．ワードサラダはスパムメールやスパムブログの大量投稿に用いられる手法であり，Web から収集した文章から作成した N-gram 頻度データを基に n 階マルコフ連鎖により確率的に文を合成する．ワードサラダはコーパスの特徴を反映した文法的に正しい文章を合成するが，合成された文章は人が見れば話題の繋がり方等から不自然であると容易に判断出来る．文法的には正しい為コンピュータには判別が困難である．

提案手法では自然な文とする文章とワードサラダを合成するためのコーパスとなる文章を収集し，コーパスから合成したワードサラダと自然な文をランダムに一つずつ提示し，設定された閾値以上の精度で正しく答えられるか否かで人と不正プログラムを判別する．CAPTCHA には，(1) 問題の合成が機械的に可能であること，(2) 不正プログラムには判別が困難であること，(3) 低賃金労働者による攻撃に頑強であることといういくつかの必要条件がある．ワードサラダは，１つのコーパスから異なる文を大量に合成出来る為，条件 (1) を満たしている．提案手法では文の不自然さを理解出来るネイティブレベルの言語能力

[†1] 東海大学大学院工学研究科

が必要となる為，外国人低賃金労働者を用いたリレーアタックに対しても耐性があり，条件 (3) も満たしていると期待される．

しかしながら，条件 (2) の不正プログラムには，単純な総当り攻撃だけでなく，自然言語の知識ベースによる文章校正機能による攻撃も考えられる．また，条件 (3) についても，日本語以外の言語についてもワードサラダが有効であるかどうか不明である．言語に応じた合成方式を検討する必要がある．

そこで，本稿では，ワードサラダが Microsoft Word の文章校正ツールによって校正されるときに不正プログラムによって判別出来ると仮定し，その安全性を厳密に評価する．日本語以外に英語，中国語，タイ語での提案手法の実装を行い，その有効性から提案手法の他言語への適用可能性について検討する．以上の評価結果に基づき，機械が一定の確率で問題が検知出来る時に，選択式の CAPTCHA において最も効率良く問題を解くためのアルゴリズムを改良する．そして，それを考慮した場合の提案手法の最適な閾値と，それについての精度とパフォーマンスについて評価する．

2. n 階マルコフ連鎖による文章合成のアルゴリズム

n 階マルコフ連鎖による文章合成は，コーパスから抽出した n-gram 頻度データに基づいてマルコフ連鎖モデルを作り，人工的な文章を合成する手法である．マルコフ連鎖による文章合成で i 番目に出力される語 x の確率は以下の条件付き確率に従う．

$$P(x_i) = P(x_i|x_{i-1}, x_{i-2}, \ldots, x_{i-N})$$

日本語は分かち書きされていないため，日本語の単語単位の n-gram 頻度を得る為には前処理として形態素評価器による分かち書きが必要である．抽出した単語 n-gram 頻度データ基に n 階マルコフ連鎖で文章を合成する．以降，マルコフ連鎖により合成された文章をワードサラダと表記し，ワードサラダ合成に用いたマルコフ連鎖の階数を階数 n とする．前述の様に n-gram 言語モデルでは，n-gram で切り出された語集合での $N-1$ 番目の語から N 番目の語を推測する事が可能である．そのため，n 階マルコフ連鎖で文章を作るためには $n+1$-gram 頻度データが必要となる．

3. 従来研究[12)]

3.1 提 案 手 法

提案手法ではコーパスから合成したワードサラダと自然な文をランダムに一つずつ提示し，設定された閾値以上の精度で「自然」か「合成」かを正しく答えられるかどうかで人と機械を判別する．h，s をそれぞれ，自然な文の数とワードサラダの数とし，$h+s$ を c と定義する．自然な文は収集したコーパスから一部の文を利用する．

認証プロセスでは，自然な文とワードサラダとを正しく解答した回数を正解数 k とし，k が閾値 θ 以上ならば CAPTCHA 成功とする．業績 1 で行った実験では，提案手法はリレーアタックや総当たり攻撃に対して耐性を持つ事を示した．しかし，CAPTCHA に掛かる時間は 308 秒となり，パフォーマンスは低い事がわかっている．

3.2 評 価 方 法

X を入力を表す確率変数，Y を出力を表す確率変数，H を人間による文章，S をスパム (機械生成の) 文章とすると，自然な文を出題して自然と回答する条件付確率は $P(Y=H|X=H)$ と表せる．自然な文章とスパム文章を出題する確率 (事前確率) はそれぞれ，

$$P(X=H) = \frac{h}{c}$$
$$P(X=Y) = \frac{s}{c} = 1 - \frac{h}{c}$$

である．従って，自然な文章とスパム文章の歪みを考慮した CAPTCHA 成功率は，これらの同時確率 $P(X,Y)$ で次のように与えられる．

$$P(Y=H, X=H) = P(Y=H|X=H)P(X=H)$$
$$P(Y=S, X=H) = P(Y=S|X=H)P(X=H)$$
$$P(Y=H, X=S) = P(Y=H|X=S)P(X=S)$$
$$P(Y=S, X=S) = P(Y=S|X=S)P(X=S)$$

CAPTCHA の検査に失敗するには，正しい自然な文章をスパムと誤判定することとスパム文章を自然な文章と誤判定することの 2 種類があり，これらをまとめて， CAPTCHA 失敗率 P_q を以下の様に定める．

$$P_q = P(Y=S, X=H) + P(Y=H, X=S)$$

ここで，人間が CAPTCHA を試行したのに $k<\theta$ となる確率を，人間拒否率 FRR(False human Rejection Rate) と定める．また，機械による CAPTCHA 成功率を P_m とし，機械による攻撃が $k \geq \theta$ を満たす確率である機械受け入れ率を FAR(False machine Acception Rate) と定める．この時，FRR 及び FAR は確率 P_q 及び P_m の二項分布で表すことができる．

$$FRR = \sum_{k=\theta}^{c} \binom{c}{k} {P_q}^k (1-P_q)^{c-k}$$

また，$FRR = FAR$ となる値を EER(Equal Error Rate) とする．

3.3 過 去 実 験

情報系の学生 7 名に対し，1 文からなる評価データを $h = 5, s = 15$ の計 15 題を提示し，文章が機械的に合成されたものであるかどうかを判断させ，正答率と解答にかかる時間を計測した．評価データはコーパスから合成した $n = 1$ のワードサラダと，コーパスからの一部切抜きである．同一テーマの政治・経済のニュース記事をコーパスとした．

3.4 過去実験:実験結果

過去実験の正答率と応答時間を表 1 と表 2 にそれぞれ示す．これらの結果から正答率，応答時間共に最小となるのは $n = 1$ で合成したワードサラダである事が解る．その為，提案手法で提示するワードサラダは $n = 1$ で合成した物を用いる．

表 1 過去実験:同時に出題する n についての正答率

出題	$n = 1$	$n = 2$	$n = 3$
H	0.91	0.80	0.68
S	0.73	0.62	0.45

表 2 過去実験:同時に出題する n について応答時間 (s)

出題	$n = 1$	$n = 2$	$n = 3$
H	8.05	8.12	7.44
S	6.19	7.75	8.58

3.5 従来研究での評価

実験 1 から得られた文章量が 1 文の時の $n = 1$ の時の解答の正答率は，条件付確率 $P(Y_h|X)$ として表 3 の様に与えられた． 従来研究での評価では，機械による攻撃を総当り

表 3 文章量が 1 文の時の $n = 1$ の時の条件付確率 $P(Y_h|X)$

入力文書 \ 判別文書	$Y_h = H$	$Y_h = S$
$X = H$	0.91	0.09
$X = S$	0.27	0.73

攻撃と仮定し，その成功率を $P_b = 1/2$ とした．総当り攻撃の受け入れ率を FAR_b とし，以下の様に定めた．

$$FAR_b = \sum_{k=\theta}^{c} \binom{c}{k} (P_b)^k (1 - P_b)^{c-k}$$

階数 $n = 1, \ldots, 3$ のワードサラダにおいて，閾値 θ についての FRR と FAR_b の関係について，$h = 5, s = 15$ の場合を図 1 に示す．また,$h = 5$ の時の $s = 5, 10, 15$ のそれぞれの場合において，階数 n についての EER を図 2 に示す．

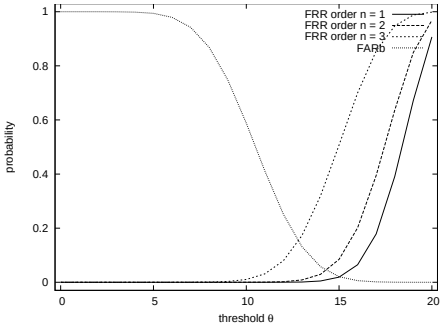

図 1 $s = 5$,$h = 15$ の時の閾値 θ についての FRR と FAR

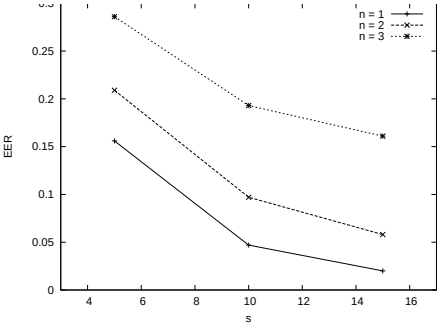

図 2 $n = 1, 2, 3$ の時の s についての EER

図 2 より，この条件では s の値が大きくなるに連れ EER の値は小さくなる事が伺える．また, 図 1 より $n = 1$,$h = 15$,$s = 5$ の場合，最も EER に近くなる θ の値は $\theta = 15$ の時である．s の値を変化させた時の FRR と FAR_b の関係を図 3 に示す．

これより，この条件で提案手法による CAPTCHA を行った場合，FRR 及び EER はお

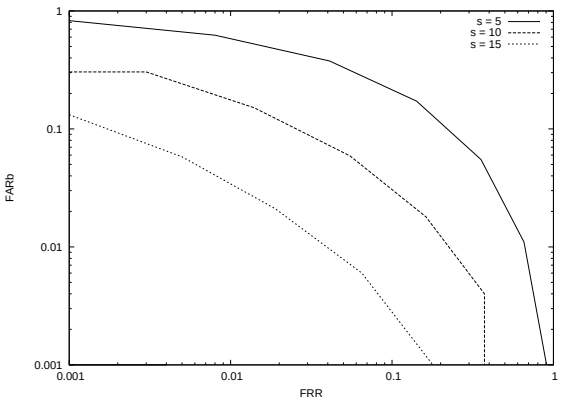

図 3 s の値を変化させた時の FRR についての FAR

よそ 3% となる事が予想される．また，過去実験の応答時間の結果から，自然な文章に対しての応答時間は 8.05 秒であり，合成された文章では 6.19 秒である．その為，この条件に於いて一回の CAPTCHA に掛かる時間は平均で凡そ 151.7 秒となる．

4. 実　験　2

4.1 実験 2:実験内容

ワードサラダは機械による自動的な判別が困難であるとされている．機械によるワードサラダ検出の精度を評価する為に，以下の内容の実験を行う．実験 2 で利用したニュースサイトの政治・経済新着記事のコーパスから，$n = 1, 2, 3$ のワードサラダと自然な文についてそれぞれ 300 文を出力し，MicrosoftWord2007 による文章校正を適用する．文中に校正箇所が示された文を検出が行われたと物して，その割合を計測した．

4.2 実験 2:実験結果

実験 2 で得られた文章校正が行われた割合を表 4 に示す． 実験では $n = 1$ の時のみ，300 件中 72 件，即ち 24% の割合で文章校正が行われた．$n = 1$ のワードサラダに置いて，校正が行われた文と行われなかった例をそれぞれ表 5 に示す．検出が行われた例では，入力ミスと判断され校正が行われた．検出が行われなかった例では，文の意味は支離滅裂だが校正は行われていない．

表 4 実験 2:文章校正が行われた割合

$n = 1$	$n = 2$	$n = 3$	自然な文
0.24	0	0	0

表 5 ワードサラダに対する文章校正の例

校正が行われた
第二次世界における影響力は、各国の影響力を樹立して いったの クリストファー・検閲等から遠洋捕鯨が 民間に送られてさらに各地から購入した。
校正が行われない
ラク・オバマ大統領の紛争や国民に対して 政治的に殆ど被害を謳歌している。

4.3 評　価　2

評価 2 では，文章校正による検出を用いた攻撃を行われた場合の提案手法の精度について検討する．提案方式について，一つの文について文章校正が行われる事象を W と定義し，その確率を $P(W)$ とする．ここで，$P(W)$ は同時確率として以下の様に表せる．

$$P(W) = P(W, X = S) + P(W, X = H)$$

この同時確率は，同時確率の定義に従い，条件付き確率として以下の様に求められる．

$$P(W) = P(W|X = S)P(X = S) + P(W|X = H)P(X = H)$$

この時，文章校正が行われた文の入力がスパムである確率は，条件付き確率で以下の様に求められる．

$$P(X = S|W) = \frac{P(W|X = S)P(X = S)}{P(W)}$$

同様に校正が行われなかった際の入力が自然な文である確率は以下の様になる．

$$P(X = H|\overline{W}) = \frac{P(\overline{W}|X = H)P(X = H)}{P(\overline{W})}$$

従って，機械は表 6 に示す文章校正を用いた検出によるスパム判定機を得る． この判定機

表 6 文章校正を用いた検出によるスパム判定機の一般式

検出 \ 入力文書	$X = H$	$X = S$
W	$P(X = H\|W)$	$P(X = S\|W)$
$\overline{W}$	$P(X = H\|\overline{W})$	$P(X = S\|\overline{W})$

を用いた場合に置ける機械の解答のアルゴリズムは以下の式で表せる．

$$Y_w = \begin{cases} S & \text{if } W = true \text{ with prob.} P(X = S|W) \\ H & \text{else if with prob.} P(X = H|W) \\ S & \text{else if } W = false \text{ with prob.} P(X = S|\overline{W}) \\ H & \text{else with prob.} P(X = H|\overline{W}) \end{cases}$$

この判定機による判定が実際に成功する確率は，それぞれの同時確率で $P(Y_w = S, X = S)$ と $P(Y_w = H, X = H)$ の二種類であり，これらは，条件付き確率として以下の様に求めた．

$$P(Y_w = S|X = S) = P(Y_w = S|W)P(W|X = S) + P(Y_w = S|\overline{W})P(\overline{W}|X = S)$$

$$P(Y_w = H|X = H) = P(Y_w = H|W)P(W|X = H) + P(Y_w = H|\overline{W})P(\overline{W}|X = H)$$

実験 2 の結果より，Microsoft Word による文章校正を 1 文ずつ行った際に文章構成が行われる確率は，条件付確率として以下の様に与えられる．

$$P(W|X = S) = 0.24$$

$$P(\overline{W}|X = H) = 0$$

これより，提案手法において 1 題あたりに校正が行われる確率は $P(W) = 0.06, P(\overline{W}) = 0.94$ となる．また，過去研究と同じ条件である $s = 15$, $h = 5$ で提案手法を行った場合，自然な文 H が出題される確率は $P(X = H) = 0.75$，不自然な文 S が出題される確率は $P(X = S) = 0.25$ となる．この時，スパム判定機を用いた攻撃では，機械は文章校正による検出が行われれば必ず「不自然」であると解答し，そうで無い場合は 60% の確率で自然を，40% の確率で不自然を選択する．この「自然」「不自然」の選択は機械による出力がそれぞれ $Y_w = H$，$Y_w = S$ となる事を表す．これらを基に，機械による攻撃の成功率は条件付確率 $P(Y_w|X)$ として表 7 の様に求められた． この機械による攻撃の成功確率を P_w と

表 7 判定機を用いた機械による条件付確率 $P(Y_w|X)$

入力文書 \ 判別文書	$Y_w = H$	$Y_w = S$
$X = H$	0.798	0.202
$X = S$	0.394	0.606

し，この判定器を用いた機械受け入れ率 FAR_w を以下の様に定義する．

$$FAR_w = \sum_{k=\theta}^{s} \binom{s}{k} (P_w)^{s-k} (1 - P_w)^k$$

$P(X = S)$ と $P(X = Y)$ の歪みを考慮すると，$P_w = 0.697$ となる．以上の値を用いて，従来研究での評価との比較を行う．

従来研究での評価の結果で使用した条件である，提案方式による CAPTCHA の問題として提示する文章量が 1 文の時，$n = 1, s = 5, h = 15, c = 20$ の場合に於いて，正解数 k の閾値 θ についての FAR_w，FAR_b と FRR を図 4 に，FRR についての FAR_b,FAR_w を図 5 にそれぞれ示す．

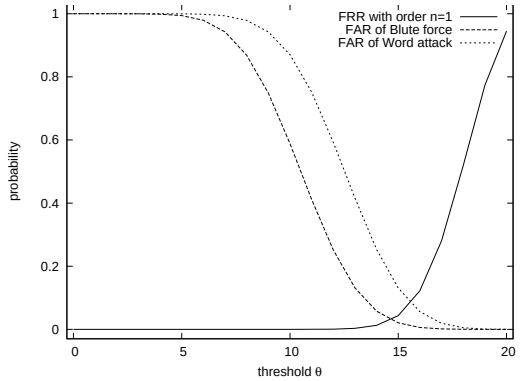

図 4 正解数 k の閾値 θ についての FAR_w，FAR と FRR

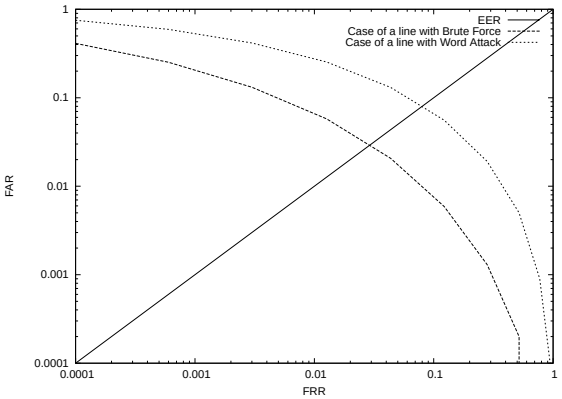

図 5 FRR についての FAR,FAR_w

以上の評価より，評価 3 と同様の $n = 1$, $s = 5$, $h = 15$, $c = 20$ の条件で提案方式を行い，機械により文章校正ツールを用いた検出が行われた時，EER は 23%程度となり，閾値 $\theta = 17$ の時 $FRR = 24\%$, $FAR_w = 22\%$ となる事が予想される．

4.4 評　価　3

評価 2 では，自然な文と不自然な文の出題数の偏りも精度の低下を招く大きな要因となった．その為，CAPTCHA として提示するの自然な文と不自然な文の出題数はあまり偏らない事が望ましいと考えられる．$c = 20$ の時，s と h の値の割合についての機械の正答率 P_w と P_r を図 6 に示す． 図 6 より，P_w が最小になるのは s と h が同数の場合ではなく，s と

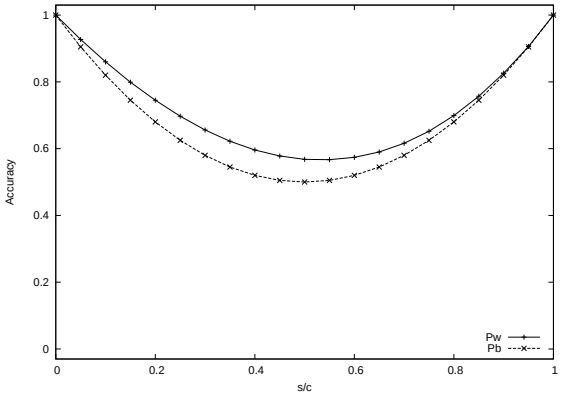

図 6　s/c についての機械の正答率

h の割合が 4.5 : 5.5 の時である事が分かる．人間の一題辺りの CAPTCHA 失敗率 P_q も問題に含まれるスパムの割合により変化する為，より適切な条件を求める．

$c = 20$ の時，$s = 9, 10, 11$ のそれぞれの場合についての FRR と FAR_w の関係を図 7 に示す． 図 7 より，$c = 20$ の場合では $s = 9$ の時に EER が最低値となった．この時の EER は 15% 程度となり，評価 2 の結果よりもおよそ 8% 改善された．

5. 提案方式の他言語への適用

マルコフ連鎖を用いて合成されるワードサラダは，コーパスについて，形態素同士の繋がりの確率情報しか必要としない．よって，ワードサラダを用いた提案方式による CAPTCHA は他言語にも適用可能であると考えられる．以下の実験で提案手法の他言語への適用可能性

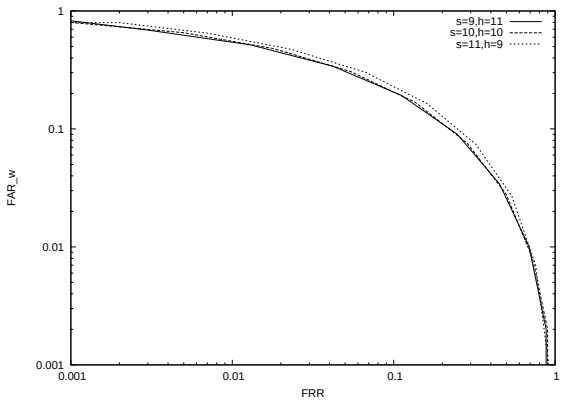

図 7　スパムの割合による ERR の変化

を検証する．

5.1 実　験　3

ワードサラダが適用可能である言語の条件を求める為，以下の実験を行った．日本人学生 3 名，英語，中国語，タイ語を母国語とする学生それぞれ 1 名に対し各言語の評価データを提示し，正答率を計測した．評価データは，文章の意味を揃える為に Wikipedia のアメリカ合衆国の記事の本文から抽出し合成した．全てのコーパスについて，ワードサラダが明らかに不自然になりすぎる要因である為，括弧表記を全て削除した．形態素解析器にはそれぞれ，日本語は MeCab，中国語は ICATALS を用いた．英語はスペースのみで分かち書きを行い，タイ語はネイティブによる手作業で単語単位に分割を行った．実験で使用した問題として提示した文は以下の様な物である．タイ語，中国語については文字コードの関係上画像で示す．

日本語

$S(n = 1)$

朝鮮戦争が積極的に起こる第二次世界大戦が完全にはニューヨークに発効されて構成され、自由のコントラ支援した。

$S(n = 3)$

大戦以前は非戦争時には GDP に対する軍事費の比率が低い国だったが、大量破壊兵器は見つからず石油を狙った侵略行為と批判する声があがった。

H

後にアメリカ人は「明白な天命」をスローガンに奥地への開拓を進め、たとえ貧民でも自らの労働で土地を得て豊かな暮らしを手に出来るという文化を形成して「自由と民主主義」理念の源流を形作っていった。

英語

$S(n=1)$

With the vast bulk of Englishmen and the Louisiana territory separated from the Southwest, which states were organized on September 17, the 100 th century, has been described.

$S(n=3)$

In 1507, German cartographer Martin Waldseemuller produced a world map on which he named lands of the Western Hemisphere America after Italian explorer and cartographer Amerigo Vespucci.

H

The United States of Americasis a federal constitutional republic comprising fifty states and a federal district.

中国語↓
S(n=1)↓
这些数据均与人类在该州获得参议院三分之二通过后，在长岛等地挑选和山谷，↓
最后终于废除了朝鲜战争后，并且对当地动植物保护区域全部加起来高达14,000英尺。
↓
S(n=3)↓
经历独立战争后，美国获取加利福尼亚州、内华达州、↓
犹他州全部地区，科罗拉多州、亚利桑那州、新墨西哥州和怀俄明州部分地区。↓
↓
H↓
1607年，位于伦敦的弗吉尼亚公司在北美切萨皮↓
克湾的詹姆斯敦建立英国的第一个短暂殖民地。←

図 8 実験 6. 中国語についての例題

タイ語

S(n=1)

การเพิ่มขึ้นรอบเกรดเลกส์และยุโรปเป็นอันดับ 3 หรือโรงเรียนรัฐบาล ได้รับการยุบสหภาพโซเวียต
ส่งผลงานและส่งเสริมชาติ ต่อมาเป็นถิ่นฐานง่ายขึ้นนี้ภาษาทางตะวันตกทำให้ทาส 7 รัฐฮาวาย
แต่ถูกบังคับบัญชาของอังกฤษ เพื่อใช้โดยทั่วไป รวมทั้งหมู่เกาะอลูเตียนในปี ค.ศ.1845
แนวคิดของผู้อพยพมาเป็นผู้อพยพยังมีการศึกษา นักเรียนสามารถเลือกในปี ค.ศ.1789

S(n=3)

หรือมหาวิทยาลัยเอกชน โดยนักเรียนสามารถกู้เงินจากทางธนาคารหรือหน่วยงานราชการสำ
หรับจ่ายเป็นค่าเล่าเรียนในระดับนี้ และจ่ายคืนภายหลังจบการศึกษา
มหาวิทยาลัยเอกชนส่วนใหญ่ค่าเรียนจะแพงกว่ามหาวิทยาลัยรัฐ
นอกจากนี้ภาษาที่มีใช้กันมากในสหรัฐอเมริกามากกว่าหนึ่งล้านคน
ได้แก่ ภาษาสเปน ภาษาจีน ภาษาฝรั่งเศส ภาษาเวียดนาม และ
ภาษาเยอรมัน

H

ความขัดแย้งระหว่างชาวอาณานิคมอเมริกันและชาวอังกฤษระหว่างยุคปฏิวัติราวคริสตทศวรรษ 1760
และต้นคริสตทศวรรษ 1770 นับไปสู่สงครามประกาศอิสรภาพสหรัฐอเมริกา อันเป็นสงครามที่เกิดขึ้นระหว่างปี
ค.ศ.1775-1781
เมื่อวันที่ 14 มิถุนายน ค.ศ. 1775 รัฐสภาภาคพื้นทวีป เปิดประชุมในฟิลาเดลเฟีย และก่อตั้งกองทัพภาคพื้นทวีป
ภายใต้การบังคับบัญชาของจอร์จ วอชิงตัน การประกาศว่า "มนุษย์ทุกคนเกิดมาโดยเท่าเทียมกัน"
และมนุษย์ทุกคนมี "สิทธิซึ่งไม่อาจโอนให้กันได้อย่างแน่นอน" รัฐสภาได้ประกาศคำประกาศอิสรภาพสหรัฐอเมริกา
โดยคำร่างส่วนใหญ่เป็นผลงานของโธมัส เจฟเฟอร์สัน เมื่อวันที่ 4 กรกฎาคม ค.ศ.1776
ในวันดังกล่าวได้มีจัดการเฉลิมฉลองขึ้นทุกปีเพื่อระลึกถึงวันอิสรภาพของสหรัฐอเมริกา ในปี ค.ศ. 1777
ข้อบังคับแห่งสมาพันธรัฐได้ก่อตั้งรัฐบาลสมาพันธรัฐขึ้นอย่างหลวม ๆ ซึ่งมีอำนาจจนถึงปี ค.ศ. 1789

図 9 実験 6. タイ語についての例題

5.2 実験結果

実験 2 の結果を表 8 に示す．結果から，各言語とも階数 n の値が低い時にスパム文書に対し不自然であると感じる割合は高くなった．タイ語については，全ての場合で不自然であるという結果になった．

6. 考　　察

安全性に対する評価では，完全な総当り攻撃を想定した場合と比較して，文章公正ツールを利用した攻撃を想定した場合の EER はおよそ 12%増加し 15%となった．この値は

表 8　実験 3:各言語毎の不自然な文の判別精度

Language	$n=1$	$n=2$	$n=3$	Natural
Japanese	0.87	0.47	0.20	0.90
English	1.0	0.8	0.6	0.7
Chinese	1.0	0.8	0.5	0.7
Thai	1.0	1.0	0.8	0.6

CAPTCHA の精度としては充分な値ではあるが, 提案手法は応答時間を基準とするパフォーマンスには優れていない為，ユーザビリティの向上に向けて更なる改善が求められる．今回はワードサラダの検出を MicrosoftWord により行ったが，ワードサラダのフィルタリングに用いる手法を利用した検出を行う事も可能であると推察される．この様な様々な攻撃方法を想定する事により，従来の CAPTCHA と同様に機械による攻撃に弱いと判断された提示問題を除外する事で精度の向上を図る事が出来る．

実験 2 の結果では, タイ語のみ全ての場合で不自然であるという結果になった．この原因は形態素解析を手動で行った為であると考えられるが，タイ語には文末の記号が存在せず，更に分かち書きされていない言語であり，形態素解析は困難な言語である．日本語，中国語，英語については各言語とも同じ様な振る舞いをする結果が得られた．各々文法規則や分かち書きの有無等大きく異なる言語である為，適切な形態素解析さえ行えれば提案手法は他言語にも適用可能であると考えられる．

7. 結　　論

本稿では，合成された文章の不自然さを利用した CAPTCHA について更なる安全性評価を行い，提案手法の改善を検討した．また，提案手法の日本語以外の適用について検討し，適切な形態素解析がその条件であると言う見込みを得た．問題提示方法や文章合成手法の調整によるパフォーマンスの向上と，他言語への適用についての再実験を今後の課題とする．

参　考　文　献

1) The Official CAPTCHA Site, (http: www.captcha.net)
2) J. Yan and A. S. E. Ahmad: Breaking Visual CAPTCHAs with Naive Pattern Recognition Algorithms, 2007 Computer Security Applications Conference, pp. 279-291，2007.
3) J. Elson，J. Douceur，J. Howell and J. Saul，Asirra: a CAPTCHA that exploit interest-aligned manual image categorization, 2007 ACM CSS, pp. 366-374, 2007.
4) P. Golle: Machine Learning Attacks Against the ASIRRA CAPTCHA, 2008 ACM CSS, pp. 535-542 2008.
5) 山本匠，J. D. Tygar, 西垣正勝：機械翻訳の違和感を用いた CAPTCHA の提案，情報処理学会研究報告, CSEC-46 No. 37，2009.
6) 山本匠，J. D. Tygar, 西垣正勝：機械翻訳 CAPTCHA(その 2)，コンピュータセキュリティシンポジウム 2009 論文集，pp. 211-216 (2009.10)
7) 鈴木徳一郎，山本匠，西垣正勝：リレーアタックに耐性をもつ CAPTCHA の提案，情報処理学会研究報告, CSEC-48 No. 16, 2010.
8) T. Larvergne, et al.,：Detecting Fake Content with Relatine Entropy Scoring', CEVR, Vol.377, pp. 27-31, 2008.
9) 森本，片瀬，山名：N-gram と離散型共起表現を用いたワードサラダ型スパム検出手法の提案, 情報処理学会研究報告, DBS-148, No.24, pp.1-8,2009.
10) 鴨志田芳典，菊池浩明：マルコフチェインによるワードスパムの合成実験とその評価について，第 72 回情報処理学会全国大会，講演番号 2G-1，2010.
11) MeCab，MeCab: Yet Another Part-of-Speech and Morphological Analyzer, (http: mecab.sourceforge.net/)
12) ・鴨志田芳典，菊池浩明, “ 文章合成の不自然さの評価と応用 ”，第 26 回ファジィシステムシンポジウム (FSS2010)，pp. 1069-1074，2010.