

印象に残りやすい日本語パスワードの合成法について

佐藤 智貴[†] 長谷川 真也[†] 鴨志田 芳典^{††} 菊池 浩明^{††}[†] 東海大学情報理工学部情報メディア学科 ^{††} 東海大学大学院工学研究科情報理工学専攻

1 はじめに

パスワードは最も広く使われている利用者認証方式である。しかし、パスワードによる認証のほとんどは英数字に限定されているため、安全性を上げるために無意味で、複雑な組み合わせを覚える必要性があった。

そこで、西坂らは携帯電話を対象とした日本語パスワード方式を提案している [1]。日本語はひらがな、カタカナだけでも 100 文字以上、漢字を含めると 2000 文字以上もあり、英数字よりも総当たりが強くなり、日本人にとって覚えやすいパスワードを生成できると主張している。[1] では、T9 入力方式利用し、ランダムな PIN を日本語に変換している。しかしながら機械的に生成されたパスワードは必ずしも覚えやすいとは限らない。

そこで本研究では、2 つの単語の合成語の Web での出現頻度に注目する。合成することにより、相殺効果を生じ、極端に頻度を下げようとする 2 語は、印象に残りやすいパスワードであると仮定する。その信頼性を検証するために、Google N-gram データベース [2] を用いて、実験及びアンケートを行い、本方式の有効性を定量的に評価する。

2 提案方式

2.1 不整合性 Confliction

単語 1 の出現頻度を W_1 、単語 2 の出現頻度を W_2 、合成語 $x = \text{“ 単語 1 単語 2 ”}$ の出現頻度を S とする*。合成語の不整合性 $C(\text{Confliction})$ を

$$C_x = -\frac{1}{10} \log \frac{S}{W_1 + W_2} \quad (1)$$

と定める。 C が大きいほど、印象に残りやすい語と考える。

2.2 提案方式の合成方法

本研究では、印象に残りにくい単語を排除するため、Google N-gram で出現頻度の高い 10,000 語を使

用する。面白く、インパクトのある語を合成するために、1) 副詞+名詞、2) 名詞+名詞の組み合わせをランダムに作り、表 1 の様に C の大きな物を選んだ。

表 1: 合成語の例

語 1	語 2	W_1	W_2	S
プライバシー	祭り	1.39×10^7	1.17×10^7	2
革命	ばあちゃん	3.97×10^7	6.5×10^6	1
コーン	公	1.04×10^7	1.18×10^8	5
次第に	くれる	9.87×10^6	7.58×10^7	8
やっぱり	戦う	6.89×10^7	1.23×10^7	9000
ひたすら	書き込み	1.69×10^7	3.74×10^7	6630

3 実験

3.1 印象度 Impression

合成語の印象の強さは、アンケートにより、5 段階で評価した。被験者 j の平均を \bar{I}_j とする。語 x の j による印象度を $I_{x,j}$ とすると、 n 人の平均を

$$I_x = \frac{1}{n} \sum_j I_{x,j} - \bar{I}_j \quad (2)$$

と定める。

3.2 実験方法

提案方式で合成した表 1 の語全 16 個について、被験者 16 名に、一人 4 つずつ覚えてもらい、3 日後にテストを行った。

3.2.1 精度 Accuracy

被験者 j は、語 x について 3 回まで試行する。1 回目で正答したら $a_j=3$ 、2 回目目で $a_j=2$ 、3 回目目で $a_j=1$ 、不正解を $a_j=0$ とする。語 x の精度を

$$A_x = \frac{1}{3n} \sum_j a_{j,x} \quad (3)$$

と定める。

3.3 実験結果

C 、 I 、 A の関係を図 1、2、3 に各々示す。

3.4 考察

図 1 の、 C の値と印象度 I の相関係数は 0.617 であり、正の相関が見られる。 C の値が大きいほど、印象に残りやすい。

Synthesis method of impressive Japanese password

[†] Tomoki Sato, Shinya Hasegawa (kakiyan.gokanoamou@cs.dm.u-tokai.ac.jp)

School of Information Science Technology and Electronics, Department of Information Media Technology, Tokai University

^{††} Yoshifumi Kamoshida, Hiroaki Kikuchi (syake@cs.dm.u-tokai.ac.jp, kikn@tokai.ac.jp)

Graduate School of Engineering, Course of Information Science and Engineering, Tokai University

* ただし、検索エンジンで形態素に分解されない様に、“単語 1 単語 2”と完全一致を指定する。

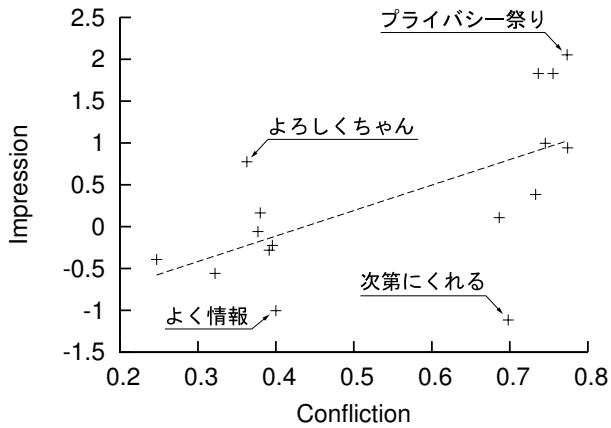


図 1: C についての印象度 I

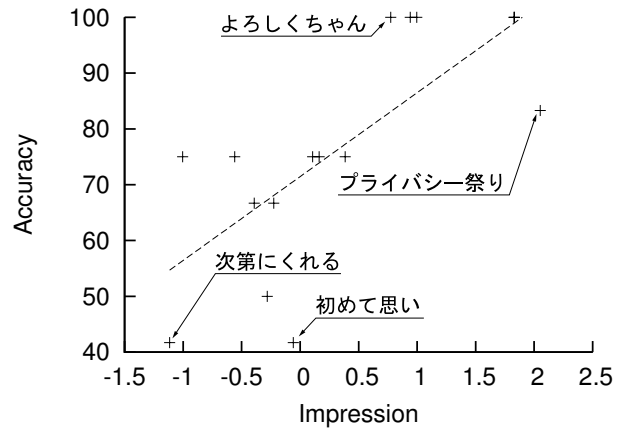


図 3: 印象度 I についての正解率 A

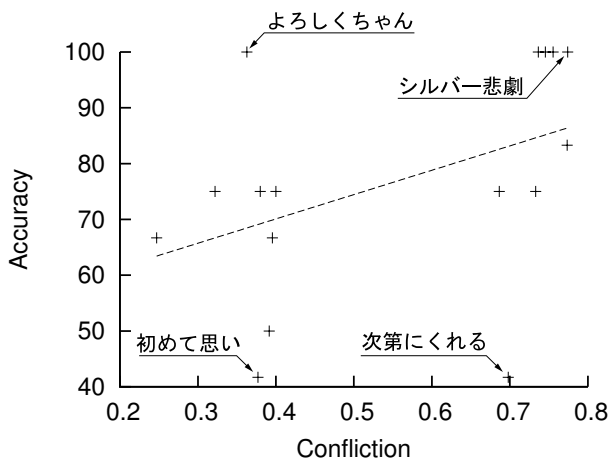


図 2: C についての正解率 A

図 2 の相関係数は、0.431 であり、正の相関が見える。 C の値が大きいほど、精度が高い。間違え方の例を表 2 に示す。“ やっぱりおかしい ”のように、片方の語は覚えているが、もう片方の語は意味や語感の似ている別の言葉を答えてしまっているケースや、“ 初めて思い ”のように、本来は漢字の部分がひらがなになっていたり、同音の別の漢字になっているケースが見られた。

“ 初めて思い ”は C の値は高いが、精度は低い。これは、良く使われる変換候補が複数存在したため、変換を間違えてしまったと思われる。図 2 で特異な“ よろしくちゃん ”は、とあるお笑い芸人がこの言葉を使用していた為、被験者の耳に残り、精度が高く、 C が低くなった可能性がある。

図 3 の相関係数は 0.733 であり、印象度と正解率の間に正の相関が見える。印象度が高ければ高いほど、精度が高い。

以上により、提案方式が有効であると言える。

表 2: 間違え方の例

実際の語	回答	間違え方
プライバシー祭り	プライベート写真	似ている言葉
やっぱりおかしい	やっぱりあやしい	似ている言葉
初めて思い	はじめて思い	変換ミス

4 おわりに

本研究では、印象に残りやすい日本語パスワードの合成法を提案し、実験及びアンケートにより評価を行った。日本語の場合は漢字の変換ミスや覚え間違い等があるので、実際にパスワードとして利用する際は、それらの点に注意する必要がある。

今後の課題としては、変換ミスや覚え間違いを防ぐ方法の考案や、パスワードの強度についての実験などが挙げられる。

参考文献

- [1] 西坂健太郎, 寺田真敏, 土居範久, “ 携帯電話を対象とした PIN 認証向け日本語パスワードの提案 ”, 情報処理学会 コンピュータセキュリティ 研究報告, Vol.2009-CSEC-048, pp.1-8, (Mar, 2010) .
- [2] Google Japan Blog : 大規模日本語 n-gram データの公開, (<http://googlejapan.blogspot.com/2007/11/n-gram.html>) .