

遺伝的アルゴリズムを用いた合理的なキーボード配列の探索

情報班: 市川弦慈、坂田響

要約

本研究では遺伝的アルゴリズムを用いて合理的なキーボード配列を探索し、新たな配列を提案した。

1. はじめに

現在のキーボード配列はタイプライターの発明に遡る歴史的な変遷を経て形作られたものなので、それが日本語入力に特化しているとは言い難い。そこで日本語入力に特化した配列を探索することにした。この研究により初学者の技術習得を容易にし、デジタルデバイドの緩和・解消を目指す。

2. 方法

(1) 概要

本研究では、遺伝的アルゴリズム(Genetic Algorithm、以下GA)を採用した。GAは最適化手法の1つであり、“遺伝子”と呼ばれる解の候補を、を“進化”すなわち徐々に最適化させることによって最適解を見つけるものである。本研究では、遺伝子を生成、評価、進化というステップのプログラムをプログラミング言語の一つであるPythonを用いて構築した(図2-1-1)。GAを利用するにあたって、遺伝子の定義、評価方法および遺伝子の進化の設定が必要になるが、本研究においては、キーボード配列をアルファベット26文字および長音記号の計27文字を文字列として遺伝子とする。(図2-1-2)。

全体像

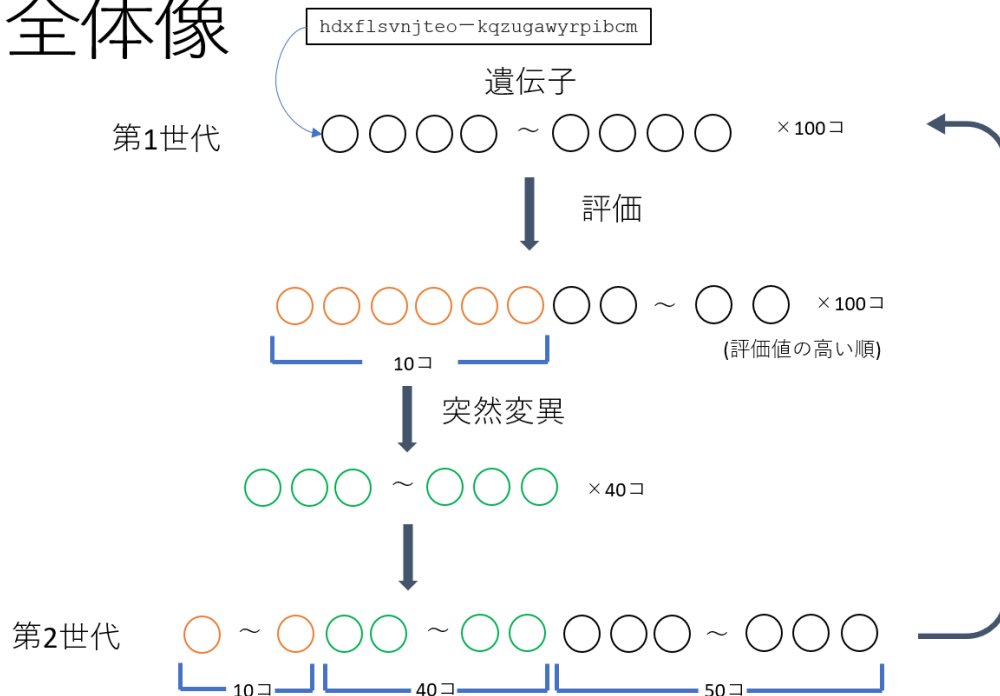


図2-1-1

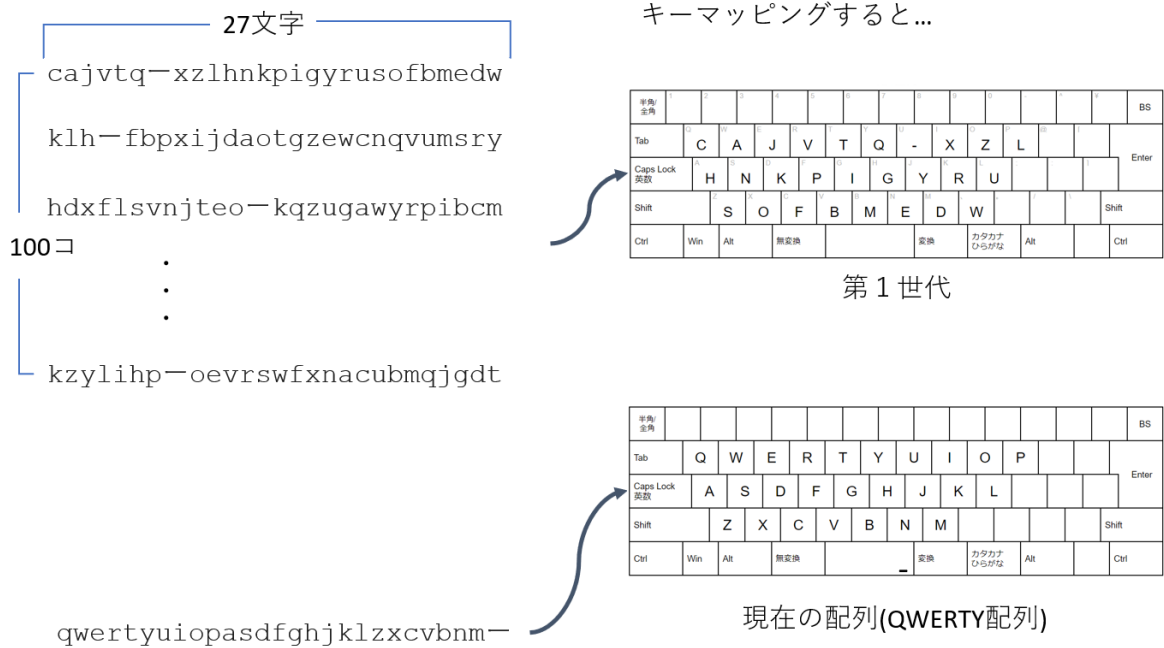


図2-1-2

(2) 評価方法

各キーにホームポジションからの入力のしやすさに応じて配点した。入力の容易なものは小さな値を、反対に困難なものには大きな値を付けた(図2-2)(以下配点表)。この配点表に従い、日本語の文章を入力するときに使われるキーの値の合計が評価値となる。(したがって、評価値が小さいほど良い解といえる。)

	Q	W	E	R	T	Y	U	I	O	P	Ⓒ
	4	2	2	3	4	5	3	2	2	4	
ick	A	S	D	F	G	H	J	K	L	:	
	1.5	1	1	1	3	3	1	1	1		
	Z	X	C	V	B	N	M	,	.	f	/
	4	4	3	2	5	3	2	3	4		

図2-2

(3) 評価データの準備

本研究ではローマ字入力を想定しているため、日本語の文章をローマ字表記したものが必要となる。しかし、一般的に日本語の文章にはかな文字と漢字が混在している。そこで本研究ではまず、任意の日本語の文章をローマ字表記に変換する独自のプログラムを開発した。これは2つの段階に分かれている。まずはMecabという日本語形態素解析ライブラリで全文章をカタカナに変換し、その後カタカナ表記とローマ字表記を対応させる。

(4) 遺伝子の進化

評価値の高かったものはその性質を次世代に引き継ぎ進化させる。本研究では”突然変異”という遺伝子のランダムな変更操作を用いた。これは、評価値の高い遺伝子を取り出し、その文字列の一部を入

れ替えるというプログラムである(図2-4)。

改良前

rlqgkhwyjpoctnv — uaizedbfsxm

改良後

rlkgqvwyjpoctnh — uaizedbfsxm

図2-4

3. 実験

(1) 目的と方法

提案手法の有効性を確認することを目的としてキーボード配列の最適化実験を行った。本研究において遺伝子の評価のために日本語版Wikipedia(テーマ:イギリス、文字数:約6000文字、QWERTY配列での評価値:1195.0)を採用した。具体的な数値に関して、第1世代の遺伝子を100個体用意し、第2世代には評価値の高かったもの50個体、突然変異させたものを40個体、新たにランダムに生成した遺伝子を追加した。同様の手順で100世代まで繰り返す。

(2) 結果

QWERTY配列と比較して評価値の推移を見ていく。このグラフは、縦軸が世代中最小の評価値、横軸が世代数である。最初は評価値1189.0であった遺伝子が世代を経るごとに評価値が小さくなっている(図3-2-2)。また29世代以降、評価値が821.5から変化しなかったのを準最適解とした。

Change in evaluation value

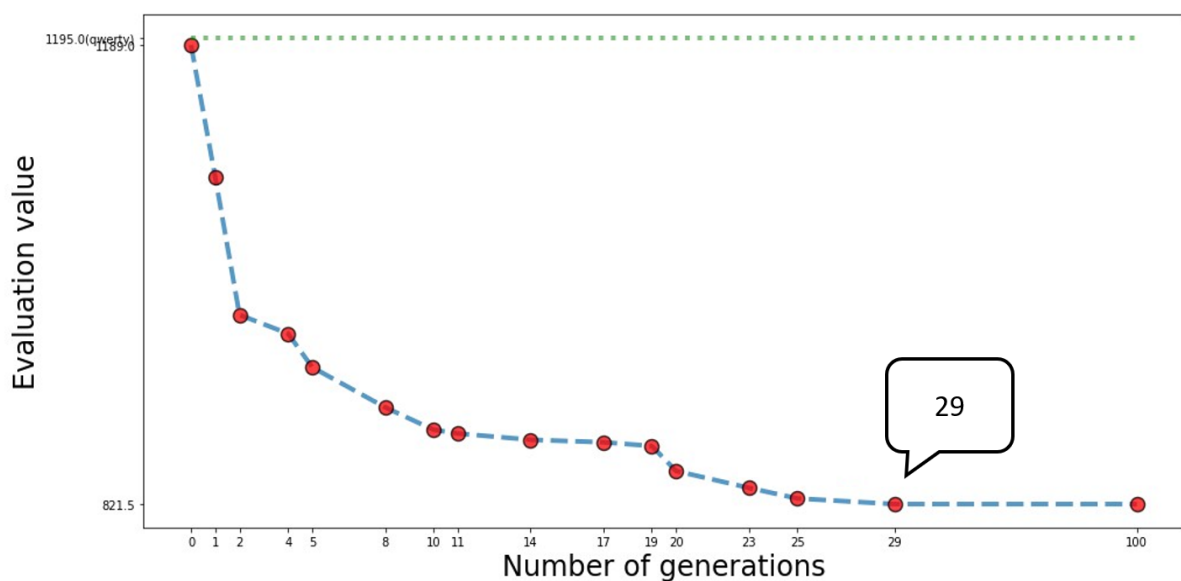


図3-2-2

4. 考察

入力頻度の高いキーが入力しやすい場所、すなわち、ホームポジションに近い位置にあるものが良い配列という評価方法であり、図4の配列では、実際に、日本語によく使用される母音が比較的ホームポジションに近い位置にある。

唯一「E」だけがホームポジションの段から離れた段にあるが、配点表で2点が与えられている場所に位置する(図2-2)。日本語入力にはあまり用いられないW,Q,Lなどのキーは配点の高いところにあり、この点では成功したと言える。ただし、Gキーは日本語において少なくともVキーよりは入力頻度が高いはずなので、プログラムにバグがある可能性があるが、見つからなかった。

	Q	W	E	R	T	Y	U	I	O	P	@
	G	H	S	Z	D	B	F	E	V	Q	
ck	A	S	D	F	G	H	J	K	L	:	
	K	T	A	I	P	-	U	N	O		
	Z	X	C	V	B	N	M	.		/	
	W	X	J	R	L	Y	M	C			

図4

5. 結論

今回、1文字ずつの評価であれば研究が成功したので、今後は2文字、3文字と繋がりのある評価方法を開発していくことを目標とする。今回の評価方法は1文字ずつで評価しているという側面があり、基本2キーで1音を入力する日本語の入力方法から考えると現実に即した評価方法とは言えないためである。

6. 謝辞

本研究を進めるにあたり、多くの方々のご指導を賜りました。大阪工業大学小林裕之教授には、研究の方針からプログラムの支援まで本研究を支えていただいたこと、改めて感謝申し上げます。