平成28年3月7日

第11回組み合わせゲーム・パズル研究集会

密度の高いシークワーズ問題作成アルゴリズム

鍋島貴大

電気通信大学 情報理工学部
情報・通信工学科 コンピュータサイエンスコース 4年
学籍番号 1211131
伊藤大雄研究室 所属

はじめに

シークワーズとは、盤面上から決められた
単語を探すパズルである

[単語リスト]

カニ	ネズミ
クロネコ	ミミズク
スカンク	ラッコ
スズメ	

シークワーズの例題

(シークワーズの遊び方、ルール、解き方 | WEB ニコリ
http://www.nikoli.co.jp/ja/puzzles/seek_words/ より)

はじめに

背景

- シークワーズを解くのはクラスPだが、問題を作るのは
  NP完全である (Hiro Ito and Shinnosuke Seki, ISORA2015)
- 他のペンシルパズルの問題作成手法は研究がなされている

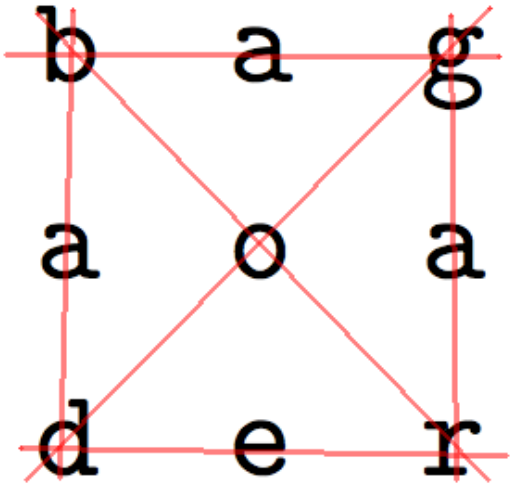
目的

- 単語のリストと盤面のサイズから、その盤面における文字の密度の上界を求める
- 上界にできる限り近い密度をもつシークワーズの盤面を作成するアルゴリズムを開発する

密度の上界

定義

n × n の盤面において、N = n × n を盤面のサイズとする。
盤面上で単語を構成する文字の総数(重複を含む)を c としたとき、c/N を、その盤面の密度とする。



bag, red,
bad, rag, → 18文字
rod, dog

密度: 18 / 9 = 2

密度の上界

使用する文字の種類をαとする

a	b	c
d	e	f
g	h	i

- 1文字の単語

  $\alpha \leqq N$ のときはα個
  $\alpha > N$のときはN個

- k文字の単語 $(2 \leqq k \leqq n)$

  縦および横方向
  n-k+1個の単語がn列ずつ

  斜め方向
  n-k+1個の単語がn-k+1列ずつ
  反対側にも同じ数

$$2\sum_{k=2}^{h} k\left\{2n\left(n-k+1\right)+2\left(n-k+1\right)^{2}\right\}$$

密度の上界

密度の上界は、αの値に応じて次のようになる

- $\alpha \leqq N$

$$\frac{c}{N} \leq \frac{1}{n^2}\left\{(\alpha + n(n-1)\left(n^2 + \frac{13}{3}n - \frac{2}{3}\right)\right\}$$

- $\alpha > N$

$$\frac{c}{N} \leq \frac{1}{3n}\left(3n^3 + 10n^2 - 12n + 2\right)$$

提案するアルゴリズム

単語の合成によるリストの作成

盤面への入力

単語の合成

リスト内の単語同士を合成し、密度の高い単語を作成する

- リストの単語は、自身に合成された単語、合成した単語も含めた文字列長、合成された単語の数 を要素としてもつ
- ある単語wが別の単語w'の一部と一致する場合、w'にwの文字数を加え、単語数を増加させる
- すべての単語の合成が完了した後、1度でも別の単語に合成した単語はリストから削除し、単語数の降順にリストを並び替える

例)

arm, army, harm

army	harm
[army, arm]	[harm, arm]
文字数: 7	文字数: 7
単語数: 2	単語数: 2

単語の入力

リストの先頭の単語から順番に入力する

- 盤面を左上から見ていき、その位置を始点として単語が入力できるか確認する
- 確認した位置に何も入力されていない、もしくは入力したい単語の先頭または末尾の文字と一致する場合、入力できる方向があるか確認する
- 入力できる方向がある場合、その方向に単語を入力する
- 単語を1つ入力するたび、入力した単語に合成されていた単語をリスト内の別の単語すべてから削除し、再びリストを並び替える

単語の入力

例) army (arm を合成済み) を入力

u	a	s	t
			a
n	m	e	c
o		v	x

‘u’が入力されているので入力不可

u	a	s	t
	r		a
n	m	e	c
o	y	v	x

方向の確認

下方向へ入力

単語の入力

例) army (arm を合成済) を入力

harm

[harm]

文字数: 4

単語数: 1

harmからarmを削除し、文字数と単語数を減らす

どこにも入力できなかった場合には次の単語に移る
すべての単語の入力を検討し終えたらアルゴリズムを停止する

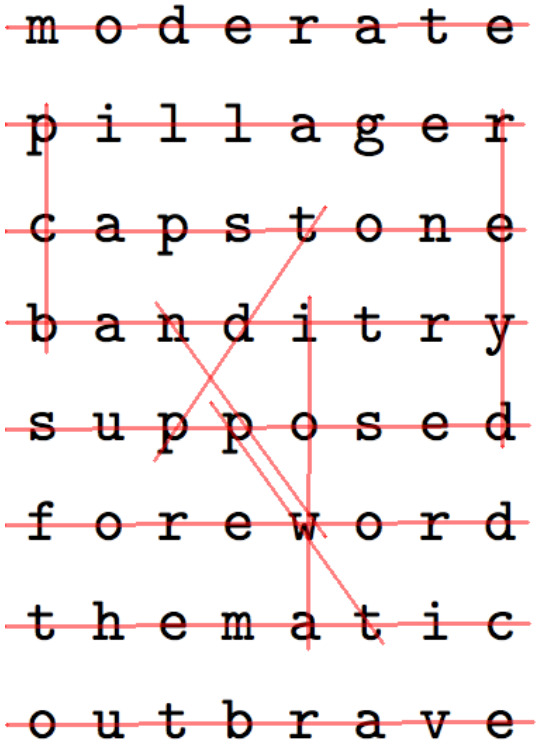
実験

アルゴリズムをPythonで作成し、計算機上で実験を行った

- 盤面のサイズを様々な値に設定し、それぞれのサイズの盤面を作成して密度を求めた
- リストには辞書の見出語を用いた
- 見出語のうち、記号や数字を含むものは除いた
- 見出語が日本語の場合、見出語の読みをひらがなで表して用いた

実験結果 1

- 英和辞書(約3万語)を用いた場合
  (くじらはんど http://kujirahand.com/ web- tools/EJDictFreeDL.php)



8 × 8 の盤面

a, ab, ac, ag, age, al, all, am, an, and, ap, ar, are, at, ate, av, ave, aw, b, ba, ban, band, bandit, banditry, br, bra, brave, bt, btu, c, ca, cap, caps, capstone, cb, ci, cit, cp, d, de, der, des, di, dn, dna, do, dom, dr, dy, dye, dyer, e, eg, eh, em, en, er, era, es, et, eta, ev, eva, ew, f, fo, for, fore, foreword, g, ga, gal, gall, ge, ger, h, he, hem, ht, i, id, il, ill, iowa, ip, it, ita, l, la, lag, lager, li, lip, ll, m, ma, mat, me, mo, mod, mode, moderate, n, na, nab, nd, ne, no, not, np, o, od, ode, of, om, on, one, op, opp, or, ord, ore, os, ot, out, outbrave, ow, owe, p, pa, pac, pc, pcb, pd, pdt, pi, pill, pillage, pillager, po, pos, pose, pp, ps, pst, pu, pus, pw, pwt, r, ra, rat, rate, rave, rb, rd, re, red, redo, reg, regal, reword, row, rower, rt, ry, s, se, so, sop, sp, spa, st, stone, sup, supp, suppose, supposed, t, ta, tam, tame, tar, tare, tb, td, te, th, the, them, thematic, ti, tic, to, ton, tone, tr, try, tsp, tu, twp, u, up, us, ut, v, va, var, w, wa, we, wo, word, wp, wpn, wt, y, yd, ye, yr

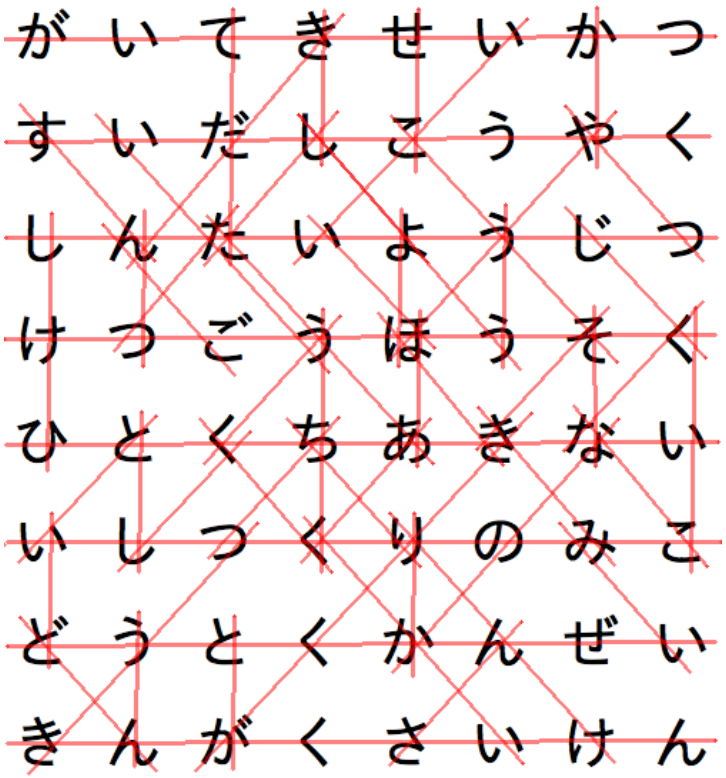
単語数　220
文字数　598
密度　　9.343750

各サイズの盤面における密度

サイズ	単語数	文字数	密度	密度の上界
2×2	12	20	5.000000	7.000000
3×3	35	69	7.666667	15.222222
4×4	63	145	9.062500	25.500000
5×5	91	213	8.520000	37.800000
6×6	133	337	9.361111	51.833333
7×7	171	457	9.326531	67.959184
8×8	220	598	9.343750	86.156250
9×9	237	660	8.148148	106.395062
10×10	298	893	8.930000	128.660000

実験結果 2

- 広辞苑(約17万語)を用いた場合
  （新村出(編): 広辞苑 第六版 DVD-ROM 版, 岩波書店, 2008)



8×8の盤面

使用した単語

あ, あうた, あき, あきな, あきない, あく, あち, あほ, い, いか, いかく, いかつ, いが, いく, いけ, いけん, いこ, いこい, いこい}, いさ, いさく, いし, いしつ, いしつくりのみこ, いす, いせ, いせき, いぜん, いた, いたん, いたんし, いだ, いだし, いて, いてき, いと, いど, いな, いよ, いよう, う, うう, うき, うこ, うご, うごつ, うじ, うじつ, うそ, うそく, うた, うたい, うち, うちく, うつ, うと, うとく, うど, うほ, うほう, うや, うやく, うよ, うん, か, かい, かいせき, かく, かくと, かくとう, かつ, かの, かや, かり, かん, かんぜ, かんぜい, が, がい, がいてき, がいてきせいかつ, がく, がくさい, がくり, がくりき, がと, がん, がんき, き, きあ, きう, きうつ, きせ, きせい, きだ, きだん, きて, きてい, きな, きない, きみ, きり, きりく, きん, きんが, きんがく, きんがくさいけん, く, くい, くいこ, くう, くか, くかい, くかん, くが, くがん, くく, くさ, くさい, くじ, くそ, くそう, くち, くちあ, くつ, くつし, くと, くとう, くな, くや, くり, くりき, くりの, け, けい, けいさ, けいさく, けし, けつ, けつご, けつごう, けつごうほうそく, けん, けんり, こ, こい, こう, こうそ, こうや, こうやく, こし, こしだ, こしだい, こせ, こな, こみ, こみの, ご, ごう, ごうほう, ごつ, ごん, さ, さい, さいけん, さく, さくが, さくがんき, さん, し, しい, しき, しく, しけ, しこ, しこう, した, しだ, しだい, しつ, しつく, しよ, しよう, しん, しんたい, しんたいよう, しんたいようじ, しんたいようじつ, じ, じう, じつ, す, すい, すいだ, すいだし, すいだしこうやく, すん, せ, せい, せいか, せいかつ, せき, せきて, せきてい, せこ, ぜ, ぜい, ぜん, ぜんか, ぜんかく, そ, そう, そうこ, そうほ, そうほう, そき, そく, そな, た, たい, たいよ, たいよう, たいようじ, たいようじつ, たう, たし, ただ, ただて, たつ, たん, たんし, だ, だい, だいす, だき, だし, だしこ, だて, だん, ち, ちく, ちくと, ちつ, ちり, つ, つう, つか, つかい, つく, つくり, つけ, つごう, つし, つじ, つち, つや, つん, て, てい, ていが, てき, てきせい, てだ, と, とい, とう, とうど, とが, とく, とくち, とし, とひ, ど, どい, どう, どうと, どうとく, どうとくかん, どうとくかんぜい, どん, な, ない, なき, なきあ, なく, なの, なのか, の, のみ, のみこ, のり, のりく, ひ, ひけ, ひけし, ひと, ひとく, ひとくち, ひとくちあきない, ほ, ほい, ほう, ほうご, ほうそ, ほうそく, ほうや, ほき, ほよ, み, みい, みき, みこ, みの, みのり, や, やか, やく, やつ, よ, よい, よう, ようじ, よし, り, りか, りき, りくつ, りち, りん, りんけ, ん, んす

単語数　332
文字数　842

各サイズの盤面における密度

サイズ	単語数	文字数	密度	密度の上界
2×2	12	20	5.000000	7.000000
3×3	42	82	9.111111	15.222222
4×4	81	172	10.750000	25.500000
5×5	140	325	13.000000	37.800000
6×6	221	522	14.500000	52.111111
7×7	287	699	14.265306	68.428571
8×8	332	842	13.156525	86.739583

2×2の盤面の総数

2×2の盤面のうち、密度が最大の7.0になるものを数え上げる

- リストから、1文字の単語をすべて抜き出す
- そのうちから4つ選んで、その4つの文字だけからなる2文字の単語のうち、合成した単語の数が最大となるものを同様にすべて抜き出し、リストとする
- 選んだ4つの文字をそれぞれ配置した盤面を作成し、2文字のリストの単語がどのように配置されるか確かめ、その密度を求める

2×2の盤面の特徴

- どの文字も盤面上の他のすべての文字と隣り合うので、文字の位置を入れ替えても盤面に配置できる単語は変わらない　⇒　<u>密度も変わらない</u>
- 文字を入れ替えてできる盤面の総数は3通り

a	b
c	d

ある1文字に注目し、その縦と横に隣り合う文字が決まれば盤面が決まる

$$ {}_3C_2 = 3 $$

2×2の盤面の総数

- 英和辞書の場合

  それぞれの文字の組み合わせに対するリストを用いてアルゴリズムを適用した結果、58通りの組み合わせの密度が最大となった (それぞれの組み合わせに対して盤面は1つずつ)

  ↓

  $58 \times 3 = 174$ 個の密度が最大の盤面を求められた

各サイズの盤面における密度（英和辞書）

サイズ	単語数	文字数	密度	密度の上界
2×2	16	28	7.000000	7.000000
3×3	35	69	7.666667	15.222222
4×4	63	145	9.062500	25.500000
5×5	91	213	8.520000	37.800000
6×6	133	337	9.361111	51.833333
7×7	171	457	9.326531	67.959184
8×8	220	598	9.343750	86.156250
9×9	237	660	8.148148	106.395062
10×10	298	893	8.930000	128.660000

2×2の盤面の総数

- 広辞苑の場合

  それぞれの文字の組み合わせに対するリストを用いてアルゴリズムを適用した結果、15855通りの組み合わせの密度が最大となった (それぞれの組み合わせに対して盤面は1つずつ)

  ↓

  $15855 \times 3 = 47655$ 個の密度が最大の盤面を求められた

各サイズの盤面における密度 (広辞苑)

サイズ	単語数	文字数	密度	密度の上界
2×2	16	28	7.000000	7.000000
3×3	42	82	9.111111	15.222222
4×4	81	172	10.750000	25.500000
5×5	140	325	13.000000	37.800000
6×6	221	522	14.500000	52.111111
7×7	287	699	14.265306	68.428571
8×8	332	842	13.156525	86.739583

まとめと今後の課題

まとめ

- 盤面のサイズから、密度の上界を計算で求めた
- リスト内の単語を合成して入力することでシークワーズの問題を作成するアルゴリズムを提案し、密度の高い盤面を作成するのに効果的であることを計算機実験で確かめた
- 2×2の盤面で密度が最大になるものを複数求めた

今後の課題

- 密度のさらなる向上のための改良
  リストの作成手法だけでなく単語の入力操作に関する研究も進める
- 密度のより厳密な上界
  文字の種類以外の条件を加える