アルゴリズム 16－3
（1990．7．16）

近傍探索法の近似度推定法

－巡回セールスマン問題を対象として－

安田 覚　阪本清和　中野秀男

大阪大学工学部

離散最適化問題に用いられる近傍探索法の良さの推定を、近似解を探索している途中で得られるデータから確率的に推定する方法について考察する。本報告では対象とする問題とその近傍探索法として、巡回セールスマン問題とλ最適法を取り上げる。５０都市程度の問題例での計算結果から、あらかじめ最適値を予測した上での推定法が近似解の出現頻度推定に有効である事を確かめた。

Estimation of Approximation for Neighborhood Search Method

for Travelling Salesman Problem

Satoru Yasuda, Kiyokazu Sakamoto, Hideo Nakano

Faculty of Engineering, Osaka University

Suita-shi, Osaka, 565, Japan

We propose a method to estimate the accuracy of local optimal solution by many data obtained during the calculation of neighborhood search method in the combinatorial optimization problem. In this report, we consider λ-opt method for travelling salesman problem. From the computational experience of the problem instances with about 50 cities, we show that it is efficient for the estimation of the frequency distribution of local optimal values to use a method conjecturing an optimum value.

1.はじめに

多項式時間では解き難い問題に対しては、一般に近似解法が用いられるが、組合せ問題やネットワーク問題のような離散的最適化問題には、近傍探索法と呼ばれる近似解法が良く用いられる。著者らは、前回の報告[1]で暗号解読問題を対象として、近傍探索法で得られる近似解の近似度の推定を行った。しかし、暗号解読問題では最適解（真の鍵）だけが求めたい解であり、近似解はすべて最適解以外は意味がないことと、暗号であるので解空間がランダムであるためもあって、近似度の推定と言うよりは何回試行すれば最適解に当たるかという推定になってしまった。

そこで本報告では、対象を比較的良く研究されている巡回セールスマン問題に取りあげ、近傍探索法も一番簡単な２本の枝の入れ替えである 2-opt法の近似度を推定する事にする。理論の構成については前回の報告から殆ど変更点はないが、この報告だけで理解できるように簡潔に前半でフローモデル等について説明する。

近傍探索法は、逐次改善法とも呼ばれ、LinとKernighanによる、グラフの２分割問題や、巡回セールスマン問題に対するλ-opt法が、近傍探索法として名前の出た始めである。この方法は、近似解法としての考え方の容易さと自然さから、いろいろな分野の問題に使われており、最近ではこの近傍探索法の一般化として、シミュレーテッド・アニーリング法も取る解の値の良さから注目されている。

近似解法の良さは、計算時間と近似度で通常は調べられるが、近傍探索法では、与えられた時間の中で多くの異なる初期解を生成し、それらから得られる局所最適解の最良値でその良さが言われる事が多い。ある問題の１つの問題例に近傍探索法を用いたとき、得られた局所解の値の分布が図1の形のように得られた時と、図2のような形が得られた時を考えると、図1の場合は経験的に最良値が最適値であると判断され、図2の場合は更に多くの異なる初期解を生成して局所解を得れば、更に良い最良解が得られると予想できる。

図1. 得られる局所最適解の分布

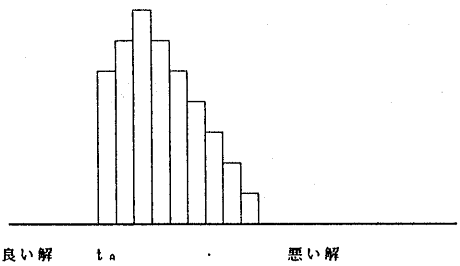
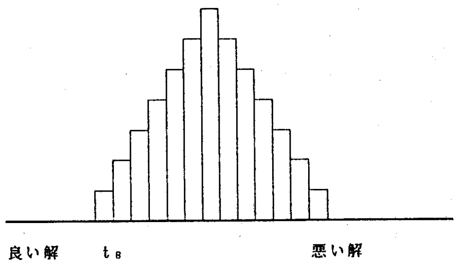


図2. 得られる局所最適解の分布



本報告では、著者らのフローモデルによる考え方をもう少し積極的に活用し、近傍探索法での計算中にデータとして得られる解の値の遷移や初期解の値の分布を活用し、それらを用いて、得られる局所解の値の分布を推定しようというものである。近傍探索法では、普通、最良解の値のみが議論の対象となり、あっても局所解の値の分布が計算時間のからみで問題となるだけであったのを、計算の途中で得られる諸々のデータをどのように活用すれば有効な近似度推定や最適解の推定が出来るかについて考察する。

また、ここでは我々の議論を確かめるために、巡回セールスマン問題に対する近傍探索

法の適用を考え、我々の提起したアプローチの是非について考察することとする。どれ程度初期解を生成したら、どの程度の近似解が得られるかを推定する。

2. 近傍探索法とフローモデル

2.1 近傍探索法

近傍探索法（Neigborhood Search Method）とは、適当に生成した初期解を解の値が改善するように解を変更しながらより良い解に近づけていく方法である。初期解の生成方法や、解の変更の等に種々の方法があるが、大略は以下の手順である。

Step.1　Step.2以下を時間の許す限り行なう。
　Step.2　初期解を生成する。
　Step.3　解を変換して値が良くなるかぎり、Step.3を繰り返す。
　Step.4　もはや解が改善されなければ、局所最適解とする。

ある問題に対する近傍探索法は、解に対する変換の集合とその変換の順序および新しい解への改善の方法を明確に与えることにより定まる。従って、ある近傍探索法が与えられた時、１つの初期解がどの局所最適解に到達するかは一意的に定まる。逆に言えば、解の集合Xはどの局所最適解に到達するかにより、各局所最適解を代表とする部分集合に分割できる。従って、初期解をランダムに与える近傍探索法では、すべての解が等確率で初期解として生成されると考えられるので、得られる局所最適解の分布は、この解集合の局所最適解による分割から分かる。ただし解空間の多峰性や解の組合せ論的な多さから、このような解集合の分割の推定は困難である。

近傍探索法では、実際には解の値が改善される方向に解が変換されていくわけであるが、解の値が改善される方向に変化（推移）していくとも考えられば、一つ一つの解を考えるよりは、値だけを考えれば良い事になり、解析がしやすいのではないかと思われる。そこで、解の集合Xをその値で分割して、X_uを値uの解の集合とし、解空間上のすべての解を個々に考えるのではなく、各X_uをそれぞれ１つの状態とし、近傍探索法を状態間の遷移としてモデル化するのがフローモデルである[2]。

以下に、フローモデルを表すのに用いられる記号を示す。

X_u　：　値uの解の集合
P_{uv}　：　X_uの解がX_vの解へ改善される確率
N_u　：　値uの初期解の個数
n_u　：　初期解が解uとなる確率
q_u　：　値uの解が局所最適解となる確率
R_u　：　値uの局所最適解の個数
r_u　：　局所最適解が値uとなる確率

初期解としてすべての解を、しかもただ１度だけ生成したとき、X_uを通過する解の個数をS_uとすると、S_uは次式で表わせる。

$$S_u = N_u + \Sigma_v S_v \ P_{vu} \tag{1}$$

ここで

$$n_u = N_u \ / \ \Sigma_v N_v \tag{2}$$

であるから、

$$s_u = n_u + \Sigma_v s_v \ P_{vu} \tag{3}$$

となる。このs_uをX_uの生起確率と呼ぶことにする。生起確率s_uとは、１回の初期解の発生および解の改善中に値uの解の表われる確率と考えればよい。

また、r_uはs_uを用いることにより次式で表わす事ができる。

$$r_u = s_u \ q_u \tag{4}$$

以上の事より、値uの解が局所最適解として得られる確率r_uは、初期解の発生確率n_uと解の遷移確率p_{uv}が分かれば求まることになる。

しかし、対象とした問題の構造が複雑であ

ったり、理論的に解析が困難であったりするとモデルとしては妥当としても、得られる最良の局所解の値の推定に使えない。また、具体的に解きたいのはある与えられた問題例であるから、その問題例に対して最良解の値や、どれほど初期解を生成したら更に良い局所最適解が得られるか推定することが必要になる。そのため、次節では近傍探索法を用いたときに得られる局所最適値の値の分布だけでなく遷移の確率等を推定する。そして、それらとフローモデルから、得られた局所最適解の値の分布だけでなく、表われてこない更に良い局所解の表われる確率を推定する方法について考察することにする。

2.2 フローモデルによる推定

2.1では、フローモデルがモデルとして妥当なら、近傍探索法のふるまいの、特に得られる局所最適解の値の分布の推定には

初期解の値の分布
解の値の遷移する確率

が分かればよいと述べた。

この2つの値は、解の値がよく表われる範囲ではデータとして集計できるし、またその値は対象とする問題例から実際に近傍探索法を試みて得られた値だけに適当な仮定をおいて理論的に導いた値より確かな値と言える。所が、値があまり出現しない範囲では、特に最適解に近づくにつれて、これらの値は推定できなくなる。初期解をランダムに生成する場合には、最適解に値が近い解が初期解として生成される確率は零ではないが無視出来るほど小さいので、この値が得られる局所最適解の値の分布の推定の計算にはほとんど寄与しないと考えてよい。

従って、フローモデルによる、得られる局所最適解の値の分布の推定では、初期解の値の分布は問題例からのデータをそのまま採用する（最良解に近い範囲では零とする）。また、解の値の遷移についても、解の値がよく表われるところではそれぞれの値についてなんらかの分布関数で近似し、値が最適解に近づくにつれてデータが少なくなるので、それまでの値での分布関数が傾向としては変わらないと予想し、それらの分布関数の最適解に近づくにつれての変化を予想することにする。このようなアプローチは妥当なように思われるが、当然、対象とする問題や問題例および採用する近傍探索法の方法によっては的確でないことも有り得る。本報告では、対象として巡回セールスマン問題を取り上げ、これに対して単純な近傍探索法の良さの推定について考察する。

3. 巡回セールマン問題とλ-opt法

3.1 巡回セールマン問題

巡回セールスマン問題(travelling salesman problem)は、古くから研究されている問題であり、一人のセールスマンが全ての都市をできるだけ早く（安く）回る巡路を見つける問題である。この問題は、いろいろな定式化が出来たり、多くの厳密解法や近似解法が提案され、その意味でも有名な問題である。

この巡回セールスマン問題は、効率の良い近似解法ではすでに数万都市の問題を３％程度の近似度で見つけだす事ができる。

3.2 巡回セールマン問題に対するλ-opt法

巡回セールスマン問題に対する近傍探索法としては、１つの可能解（巡路）に対して、それを構成している隣合わない２本の枝を、可能解を構成していない２本の枝と入れ換える2-opt法や、３本の枝を入れ換える3-opt法があり、それらを総称してλ-opt法と呼ばれるものである[3]。本報告では、近傍探索法の近似度の有効な推定法を考察する事にあるので、一番簡単な2-optについて考察する。それらの

考察から巡回セールスマン問題以外の問題や、巡回セールスマン問題でもアニーリング法等にも十分適用できるものと考えている。

4. 巡回セールマン問題に対する λ-opt法の近似度推定

4.1 局所最適値と遷移のデータ

巡回セールスマン問題に対して2-opt法を考える。例題としては、平面上にランダムに生成した50から70点程度の問題例を計算例として用いる。これは、2-opt法でも50点程度までは、最適解が得られる事が多いからであり、100点を越えると最適解から遥かに離れた解しか局所最適解として得られないからである。

フローモデルを用いた近似度の推定法では、2章で述べたように、初期解の発生確率と遷移確率が必要になる。表1は、50都市のある問題例1における解の値の遷移確率と、得られた局所解の値の分布を示したものである。値の範囲は局所解の値の範囲だけを選んでおり、その範囲内で初期解が生成される事はない。表は、解の値、局所最適解の数、その値から低い値に遷移していった解の数を示す。表1は初期解の数を1000個とした時のデータである。

表1. 50都市問題の1000回の初期解の発生

解の値	局所解の数	遷移した解の数 1	2	3	4	5	6	7	8	9	→遷移した値				
346	2	0													
347	2	0													
348	2	0													
349	1	0													
350	10	1	0	2	0	1									
351	6	3	2												
352	14	1	0	2											
353	11	0	1	1											
354	16	2	2	0	1	0	0	0	1						
355	25	2	2	2	2										
356	32	1	0	2	1	2	0	1	1						
357	20	0	4	1	1	3	0	1	1	2					
358	33	3	7	5	3	3	1	2	1	1					
359	30	5	7	3	1	5	1	3	1	2					
360	41	10	6	5	3	7	2	2	2	0	1				
361	49	11	11	11	3	8	3	1	2	2	0	0	1		
362	43	12	16	15	5	3	2	3	1	0	0	0	1		
363	57	16	18	11	3	6	2	4	1	1	1	2	1		
364	47	15	27	8	6	8	7	8	5	5	2	2	1		
365	43	13	18	12	6	7	7	6	5	3	0	1			
366	49	16	26	16	9	9	10	6	5	2	2	1			
367	56	25	19	16	8	15	6	10	4	1	2	2	1	1	1
368	50	28	29	15	13	19	5	11	4	4	1	1	1	2	1
369	38	17	29	14	15	13	6	9	9	8	3	0	1	1	
370	42	25	26	20	12	12	5	12	8	5	5	2	2	1	1
371	37	23	30	28	17	14	12	6	9	6	5	1	1		
372	21	17	36	26	14	24	11	12	5	6	4	1			
373	39	31	23	20	14	16	15	17	4	9	3	3	2	1	
374	34	41	30	37	16	18	9	12	15	6	13	4	3	0	1
375	31	40	30	31	14	18	8	23	7	10	6	5	1		
376	18	28	31	42	16	19	15	6	5	4	6	9	2	0	1
377	27	34	41	27	20	21	12	20	10	11	2	2	2	0	2
378	15	45	37	33	27	24	12	16	8	9	9	3	3	2	1

この問題例を初期解の数を10000個とした時のデータが表2である。値が360辺りからは、遷移の大きな所は省略している。

表2. 50都市問題の10000回の初期解の発生

解の値	局所解の数	遷移した解の数 1	2	3	4	5	6	7	8	9	→遷移した値		
344	1	0											
345	4	0											
346	12	0											
347	11	0											
348	26	1	1	1									
349	40	1	4	1									
350	58	4	2	3	0	3	1						
351	69	5	13	3	1	1							
352	96	10	11	8	2	3	1	2					
353	144	8	9	8	2	0	5	0	2				
354	175	19	25	5	9	2	6	0	2	1			
355	232	20	20	24	16	12	3	3	2	1			
356	272	20	18	29	13	13	6	7	6				
357	270	19	41	33	7	22	11	2	9	9	0	2	1
358	316	40	48	48	15	26	14	11	10	4			
359	352	51	56	60	16	34	19	17	11	9	2	1	
360	397	85	78	66	37	51	40	22	15	9	3	3	2
361	418	86	100	87	43	53	46	22	27	8	2	1	4
362	467	120	119	98	37	55	32	29	20	8	5	3	2
363	522	145	161	119	59	62	48	40	20	22	9	8	6
364	464	154	193	112	66	81	64	38	31	32	15	15	3
365	559	178	171	139	74	95	74	58	47	32	9	12	3
366	509	217	227	180	97	88	78	60	45	33	20	13	6
367	514	231	249	188	89	117	87	77	33	30	16	7	10
368	517	271	279	182	115	119	78	77	56	45	23	15	6
369	436	237	269	188	137	116	70	89	64	59	32	12	5
370	409	307	298	248	139	164	92	121	72	49	34	23	6
371	394	299	312	240	132	169	112	119	75	58	29	20	17
372	331	301	333	238	159	199	111	120	67	85	40	24	25
373	303	355	306	280	138	186	119	140	98	64	42	39	12
374	269	336	335	309	159	197	109	126	98	70	57	31	19
375	245	361	331	292	164	197	122	159	84	84	57	41	31
376	227	350	351	303	180	193	122	141	66	87	51	41	31
377	186	355	364	274	150	205	131	148	86	96	57	34	19
378	164	393	363	336	170	230	138	146	91	92	65	47	31

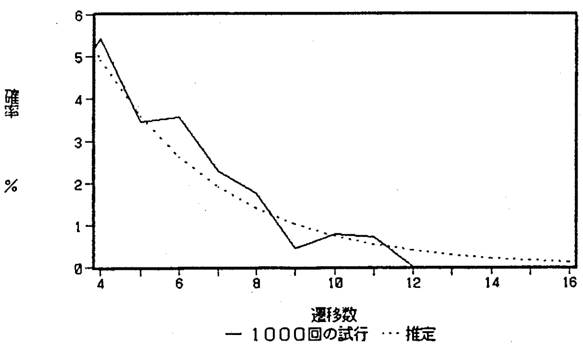
4.2 遷移確率の推定の方法

このデータから分かるように、1000回の試行より、10000回の試行のデータの方が当然データ量も多くなり、得られるデータから局所最適値の値の分布や遷移確率は推定しやすくなる。今、1000回での試行でのデータから、

遷移確率をデータが多く得られているところからは、そのデータを直接使って（但し、確率であるので正規化はする）推定し、データがあまり得られていない所や、データとして現れていないところは、それまでのデータから推定する事にする。その推定が正しいかどうかは　10000回での試行でのデータや、更にはもっと多くの試行でのデータから調べる事が出来る。もし、すべての解を初期解として一様に発生できるなら正当に評価できる事になる。

表1や2のデータでは、値が340台のデータは出現頻度が少ないので信用できないが、360台以上の値では多くのデータがあるので、そのデータをそのまま遷移確率に使う事が出来る。表1,2の各行はある値での局所最適解の数と、その値から遷移した解の数を度数分布の形で表現しているが、これを図で描いたのが図3である。ここでは良くデータが現れるところは近似のしやすい形をしているが遷移の値が大きくなると出現度数も少ないか、もしくはなくなるので推定をせざるを得ない。特に、対象とする値が最適解に近づくほどその傾向は強くなる。

図3.　遷移確率



そこで、図3の各値からの遷移確率を暗号解読の推定に使ったと同じ方法で行う。

(方法1)

各値iについて、

$$P_{ij} = C \ e^{A(i-j)} \tag{6}$$

と表し、

$$\Sigma_j \ P_{ij} = 1 \tag{7}$$

となるように回帰計算によりC、Aを定める。

このような推定法では、指数関数の性質から数値的に大きな値の遷移を小さい値ながら許す事になり現実的でない。また、実際に1000回の試行のデータから推定した値を10000回の試行のデータと比較してみても、大きな値の遷移では推定値の方が緩やかに減少するためかなりの誤差を生じる事が分かった。

そこで**方法2**として、あらかじめ最適値の値を予想し、各値から最適値までの遷移しかないものとして推定を行う。最適値までの遷移で遷移確率を打ち切る。方法1の推定式に補正を加えると考えてもよい。本報告では、この補正の所はまだ残された問題も多いので、**Lotus1-2-3**に収集されたデータを読み込ませ、作表と図示の機能を使いながら対話的に補正を行った。

4.3 計算機実験と推定結果

計算機実験は、50都市と70都市の問題についてそれぞれ２題づつ行った。実験においては問題例によっては、局所解の値の分布から見て明らかに最適解を見つけだしているときは、推定の対象からははずした。問題1の50都市の問題のデータは、表1が1000回の試行のものであり、表2が10000回の試行のものである。対象とした問題例とも、

1000回のデータを元に２つの推定法で
推定した値　　推定値1
　　　　　　　推定値2
10000回の試行での局所値の分布の値

のデータを示している。いずれの表においても解の値の一番右の欄の値は1000回の試行で得られた一番良い解の値である。

表3. 50都市の問題1に対する推定値と測定値

解の値	343	344	345	346
測定値	0	1.00×10^{-4}	4.00×10^{-4}	1.20×10^{-3}
推定値１	2.30×10^{-4}	4.38×10^{-4}	7.43×10^{-4}	2.25×10^{-3}
推定値２	0.95×10^{-4}	1.48×10^{-4}	2.43×10^{-4}	1.72×10^{-3}

表4. 50都市の問題2に対する推定値と測定値

解の値	333	334	335	336
測定値	0	0	5.00×10^{-4}	2.30×10^{-3}
推定値1	4.15×10^{-4}	6.89×10^{-4}	11.3×10^{-4}	1.82×10^{-3}
推定値2	0.12×10^{-4}	0.29×10^{-4}	0.78×10^{-4}	0.91×10^{-3}

表5. 70都市の問題3に対する推定値と測定値

解の値	405	406	407	408
測定値	1.00×10^{-4}	1.00×10^{-4}	6.00×10^{-4}	0.80×10^{-3}
推定値1	1.90×10^{-4}	3.61×10^{-4}	5.95×10^{-4}	1.95×10^{-3}
推定値2	0.30×10^{-4}	0.52×10^{-4}	0.96×10^{-4}	1.18×10^{-3}

問題4の70都市の問題では、10000回の試行でも最良解は改善されなかったので、30000回の試行も試みた。それが測定値2である。

表6. 70都市の問題4に対する推定値と測定値

解の値	382	383	384	385
測定値	0	0	0	0.20×10^{-3}
推定値1	0.23×10^{-4}	0.51×10^{-4}	0.92×10^{-4}	1.16×10^{-3}
推定値2	2.86×10^{-6}	3.11×10^{-6}	7.20×10^{-6}	1.02×10^{-3}
測定値2	0	0	3.33×10^{-5}	1.67×10^{-4}

4.4 近似度の推定法の使い方

著者らは前報告と本報告で、近傍探索法に対する近似度の推定法について考察してきた。このような推定法を実際の近似解法にどのように用いるかについてこの節では考えてみる事にする。

まず、オフライン的な使われ方で、これは得られたデータを解析し、そこから最適値を予測する事である。しかし、もともと難しい最適値の予測であるから、本推定法でもどこまで正確に予測できるかは分からないし、また本推定法がやや大まかなデータの流れから推定しているのに、最適解や最適値に非常に近い解やその周りの解空間はそれぞれの問題に合わせて特殊な形をしているかもしれない。

オンライン的な使い方とは、例えば1000回の試行でのデータから、遷移確率を推定して、10000回の試行のデータと比較して推定を是正し、更にその推定を100000回の試行のデータと比較していく方法である。表の3から6で与えた値は、逆数として見れば

何回の試行でその値の局所解が得られるか

を示しており、

あと何回試行すれば最良解が良くなるか

としても使えるものであると考えている。

推定法2では、最適値を予測して推定する方法を考えたが、オンライン的な推定法の適用では、推定の是正を最適値の予測値の是正で行うので試行を重ねるほど最適値の予測が鋭くなると考えている。

5.むすび

本報告では、近傍探索法を用いて得られる局所最適解の値の分布と、解の値の遷移確率を推定することによって、より最適値に近い局所最適解の値の分布を推定する方法を考察した。対象とする問題を巡回セールスマン問題とし、近傍探索法として2-opt法を取り上げ提案した推定法を計算機実験で検討してみた。

今回は、1000回の試行から遷移確率などを推定し、それらを10000回のデータと比較検討することで推定法の妥当性を検討したが、最適値に近い辺りでは遷移確率の推定を、Lotus-1-2-3の作表とグラフ表示の機能を使って対話的に推定値を決めた。このような推定法は、有効な推定法を見つけるまではある程度しかたがないように思える。多くの問題例に対して有効な推定のやり方が分かれば、それを計算機処理に移していく事を当面の研究方針としていくつもりである。

参考文献

[1] 中野、流谷、中西、"近傍探索法の近似度推定法に関する一考察-慣用暗号の解読を対象として-",情報処理、アルゴリズム、5-13, Jan. 1989
[2] 中野、立石、中西、"0-1ナップサック問題に対する近傍探索法のふるまいの解析"信学論、68-A、Dec, 1985
[3] Lin,S.,B.W.Kernighan,"An effective heuristic algorithm for the traveling salesman problem" Operations Research, Vol.11, No.6, Nov. 1963