2D3-2

動的環境での適応に向けた蟻コロニー最適化手法への相対評価の実装

Implementation of relative evaluation to the ant colony optimization method toward adaptation to dynamic environments

野口 直人[*1] Naoto Noguchi

高橋達二[*2] Tatsuji Takahashi

[*1] 東京電機大学大学院 Graduate School of Tokyo Denki University

[*2] 東京電機大学 Tokyo Denki University

For combinatorial optimization like Traveling salesman problem (TSP), many efficient methods have been invented. However, in a real environment, the map can change through time. A supply route for chain stores calculated as optimal can be made useless if a path is blocked up. Traffic situation that determines the weight (efficiency or distance) of paths constantly varies. We propose an algorithm which is a variant of the ant colony optimization (ACO). It implements relative evaluation of the value of paths which is known to be effective in decision-making under uncertainty.

1. はじめに

蟻コロニー最適化 (Ant Colony Optimization:以下 ACO) [M.Dorigo 96]という蟻の給餌行動を基に作られた最適化手法がある. この手法は組み合わせ最適化問題, 特に巡回セールスマン問題(Traveling Salesman Problem: TSP) [山本 97]において非常に優秀な成績を示すことが知られている[原 12]. 組み合わせ最適化問題とは条件が制約されている下で, 順序や割当のように条件を満たした解を組み合わせる問題である. この問題は多岐にわたる分野で研究されているが, 問題の規模が大きくなると解の組み合わせ数が膨大になり組合せ的爆発(combinatorial explosion) が発生する. そのため計算量が増大しやすく, 現実的な時間で最適解求めることが困難になる. しかしACO などのメタヒューリスティクス手法を用いることで, 最適解を求めることを諦める代わりに, 現実的な時間内で最適解に比較的近い準最適解を求めることができる.

組み合わせ最適化問題の例として TSP やナップサック問題などがあげられる[古川 12]. これらの問題は環境が変化しない定常環境の問題である. 他方, 現実世界では, 最適化にかかる計算時間が無視できないとすれば, 最適化の途中で環境が動的に変化する問題を扱う必要性が出てくる. ACO は探索を継続的に行うので, 焼きなまし法や遺伝的アルゴリズム[久保 09]などの, 他のメタヒューリスティクス手法に比べ非定常問題に対して有効である. しかしこの手法は, ある程度探索が進むと解の収束が起こり, 解の再探索が行われづらくなる. そのため解の収束が起こったときに環境の変化が起こると最適解を得るまでに時間がかかってしまう.

本研究では動的な意思決定課題において有用であるとされる相対評価法をACOに付加することで,既存手法において困難な動的に環境が変化する巡回セールスマン問題において有効な手法を提案する

2. 巡回セールスマン問題

巡回セールスマン問題(TSP) とは代表的な組合せ最適化問題の一つである. TSP は始めに n 個の都市集合$V = (1,2,\cdots,n)$と, 都市 i と都市 j 間のコストC_{ij}が与えられる. そして任意の都市から巡回を始め, 全都市を訪問した後に最初の都市に戻る. この巡回路の総コストが最も短い経路を求める. ここで, TSP には$C_{ij} = C_{ji}$となる対称 TSP と, $C_{ij} \neq C_{ji}$となる非対称 TSP があるが, 本論文では対称 TSP について考えた. TSP は最適解を求めようとすると, 都市数が小さい場合には容易に求めることができる. しかし, 都市数が大きくなるにつれて組合せ的爆発が起こるため経路数は飛躍的に増大し, 現実的な時間内では解けなくなってしまう. このような性質から TSP は NP（Non-deterministic Polynomial）完全問題に分類され, 都市数 n に関して, 多項式時間の解法が存在しないとされている. そこで実際的な計算時間内で, 較的良い精度の準最適解を与えるような手法が必要となっており, 昔から多くの手法が提案されてきた.

3. Ant Colony Optimization

ACO とは,蟻が巣から餌場までの経路を形成する給餌行動を基に作られた最適化手法の総称である. この手法は, フェロモンとヒューリスティック情報を基に確率的に解を生成し, 生成したその解からフェロモンを更新するというサイクルを繰り返し行う. これにより組み合わせ最適化問題の探索空間をコンパクトに限定しながら効率よく探索を行うことが出来る. この基本原理を具体化したアルゴリズムに Ant System (以下:AS) がある.

3.1 Ant System

AS とは ACO メタヒューリスティクスに基づき最初に実装された手法であり, TSP の解法として Dorigo らによって提案された手法である.AS を用い TSP を解くとき, 蟻に見立てた m 個のエージェントを各都市にランダムに配置する. 各エージェントは単純なルールのもとで「都市間に設置されたフェロモン量」と「都市間の距離の逆数」に比例した確率で都市を推移し, 巡回路を形成する.

ここで, あるエージェント, k が都市 i から都市 j に推移するときの確率p_{ij}^kは次のように定義される.

$$p_{ij}^k = \begin{cases} \dfrac{(\tau_{ij})^\alpha (\eta_{ij})^\beta}{\sum_{s\in N^k}(\tau_{ij})^\alpha(\eta_{ij})^\beta} & (if\ \ j \in N^k) \\ 0 & (otherwise) \end{cases} \tag{1}$$

ここで k はここのエージェント番号($k = 1,2,\cdots,$m)を示し, N^kはエージェント k の未訪問集合を示す.τ_{ij}は都市 i,j 間に置かれているフェロモン量. η_{ij}は問題領域固有の情報であり, TSP では一般に都市 i.j 間の距離d_{ij}の逆数を用いる. よって式(1)は, フェロモン量と問題領域固有の値に比例した選択確率が得られる. また, α と β は非負の実数で, それぞれフェロモンにより段階的に形成される大域的な情報と, それぞれ問題領域固有の値にもと

連絡先:
氏名:野口直人, E-mail:10rd178@ms.dendai.ac.jp

づく局所的な情報をどの程度重視するかを示すパラメータである.

つぎに, m 個のエージェントが式(1)を用い巡回路を生成する動作を 1 サイクルとしたとき, t サイクル目に任意の 2 つの都市 i,j 間の枝に設置されているフェロモン量$\tau_{ij}(t)$は次のように定義される.

$$\tau_{ij}(t+1) = \rho \cdot \tau_{ij}(t) + \Delta\tau_{ij} \tag{2}$$

$$\Delta\tau_{ij} = \sum_{k=1}^{m} \Delta\tau_{ij}^{k} \tag{3}$$

$$\Delta\tau_{ij}^{k} = \begin{cases} \frac{Q}{C_k} & (if\ (i,j) \in T^k) \\ 0 & (otherwise) \end{cases} \tag{4}$$

ここでρはフェロモンの蒸発率を表すパラメータであり, $0 < \rho < 1$の間で設定される. この蒸発効果により, 通らない経路のフェロモンは薄れていき, 過去の行動の情報と新しい行動の情報を適応的に変化させ解の収束を行うことが出来る. $\Delta\tau_{ij}$は, t サイクル目に m 個のエージェントが経路 i,j に設置するフェロモンの総量を示す. $\Delta\tau_{ij}^{k}$はエージェント k が都市 i,j 間に設置するフェロモン量を示しており, Q はエージェントが 1 サイクルに設置するフェロモンの定量を示すパラメータ, C_kはエージェント k が生成した巡回路の経路長であり, T^kはエージェント k が通った経路の集合である. 即ち経路に設置されるフェロモンの量は経路長の逆数によって決定されるので, 経路長が短い経路には大きい量のフェロモンが置かれ, 逆に経路長が長い経路には少ない量のフェロモンが置かれる.

AS の基本動作はまずフェロモン量の初期化を行い, 式(1)を用い m 個のエージェントが確率的に巡回路を生成する. その後式(2), 式(3), 式(4)を用いフェロモン量を更新し, 終了条件を満たすまでこのサイクルを繰り返す.

3.2 Max-Min Ant System

Max-Min Ant System(以下:MMAS) は AS を改良した手法である.AS では間違った経路にフェロモンが集中して解の収束が起こると, その経路から抜け出すことが困難になってしまう. そこで MMAS では, 各経路に置けるフェロモン量の上限値と下限値を区間 τ_{max},τ_{min}] で限定すること.さらにフェロモン量の更新に各サイクルで最も成績の良いエージェントの経路のみを用いる. 上限値と下限値で限定することにより, 解の収束が起こっても僅かな確率で全ての経路選択することが出来るので解の多様性を保つことが出来る. さらに優秀なエージェントの経路を使うことにより, 優れた経路へ素早く解の収束を行うことが出来る. MMAS ではτ_{max} , τ_{min}をそれぞれ以下の式で定めている.

$$\tau_{max}(t) = \frac{1}{1-\rho} \times \frac{1}{C_t^{best-so-far}} \tag{5}$$

$$\tau_{min}(t) = \frac{\tau_{max}(1-\sqrt[n]{P_{best}})}{(n/2-1)\sqrt[n]{P_{best}}} \tag{6}$$

ここで$C_t^{best-so-far}$は t サイクル目における最短の巡回路長であり, n は都市数, P_{best}はフェロモン量を標準化する際の下限値に関するパラメータである.

4. 相対評価

既存手法での経路の評価方法, 即ちフェロモン量の更新はエージェントが通った経路にしか設置しない絶対評価であった. しかしある経路の評価が下がった, あるいは上昇した場合, その周辺の経路に対しても価値を伝搬させ, 再評価の機会を与えることが必要だと考えられる. このような, ある評価対象の評価を他の評価対象にも影響を与える評価方法を相対評価と呼ばれ, 動的な意思決定課題においても有用であることが示されている [Tversky 74]. 相対評価を ACO に付加することにより, 解の収束による探索の安定が起こった際にも, 環境の変化を周辺の経路に伝播することで素早く環境の変化に対応することが出来ると考える.

5. 提案手法

MMAS のフェロモン更新は式(2), 式(3), 式(4)のように都市 i,j にある既存のフェロモンτ_{ij}に揮発係数ρによって減衰させ, そこに新たにエージェントの設置したフェロモン$\Delta\tau_{ij}$を加算する手法であった. この手法では, 環境の変化によって直接影響を受ける辺は, エージェントが通った辺にのみになる. そのため環境の変化にすばやく対応することが出来ないと考える. そこで本研究では, エージェントの通った経路を「選択した評価対象」, 通らなかった経路を「他の評価対象」とし, 選択した評価対象の評価を, 他の評価対象にも影響させる相対評価的なフェロモンの更新手法 Ant System with Relative Evaluation（以下:AS-RE）を提案する. AS-RE は t サイクル目で最も成績の良いエージェントが通った辺 i,k とエージェントが通っていない辺 i,j によってフェロモン更新式を以下の式(7), 式(8)に分けた.

$$\tau_{ik}(t+1) = \tau_{ik}(t) + \delta\tau_{ik}(t) \tag{7}$$

$$\tau_{ij}(t+1) = \rho\tau_{ij}(t) + \mu\delta\tau_{ik}(t) \tag{8}$$

$$\delta\tau_{ik}(t) = \rho\tau_{ik}(t) + \frac{Q}{C_t^{best-so-far}} - \tau_{ik}(t)$$

$$= (\rho-1)\tau_{ik}(t) + \frac{Q}{C_t^{best-so-far}} \tag{9}$$

$\delta\tau_{ik}$(t)は t サイクル目でのフェロモンの変化量, 既存手法での $\Delta\tau_{ij}$ に相当する. 微少係数μは辺 i,k の評価を隣接する辺 i,j に与える影響の調節を行っている. 提案手法ではエージェントが通った経路を式(7)のように辺に設置されているフェロモン量に変化量を加算する. エージェントの通らなかった経路では式(8)のように, 辺に設置されたフェロモン量から蒸発係数ρによって減少させたところに辺 i,k 間の変化量が僅かながら加算される. これは変化量$\delta\tau_{ik}$(t)が他の経路へ拡散されることに等しい. ある程度探索が進み解の収束が起こると, 経路が一定になるため式(9)の$Q/C_t^{best-so-far}$ は一定の値なるが, $(\rho-1)\tau_{ik}(t)$の値は減少していき変化量$\delta\tau_{ik}$(t)が 0 に近づき, τ_{ik}(t)が一定の値に収束する. また MMAS と同じくτ_{max}によって一定の確率で全ての経路が選択される. そのとき経路 i,j がτ_{min}の確率で選ばれ, そのτ_{ij}を含む新たな経路の総コストが小さくなる場合, 変化量$\delta\tau_{ik}$(t)が増大する. すると式(8)により他の経路のフェロモン量も増えるため, ある程度収束した後でも解の再探索を促すことが出来る.

さらに, 解の収束により変化量$\delta\tau_{ik}$(t)が 0 に近づいたとき, 環境の変化が起こりもともと通っていた経路の経路長が大きくなった場合, 変化量$\delta\tau_{ik}$(t)が負の値になり辺 i,k のフェロモン量を減少させる. 他の経路に対しても負の変化量が伝播されるが微少係数μを掛けているので辺 i,k ほど減少されない. そのことにより

相対的に他の経路の評価が上がり再探索を促すことが出来るのではないかと考えた.

6. シミュレーション

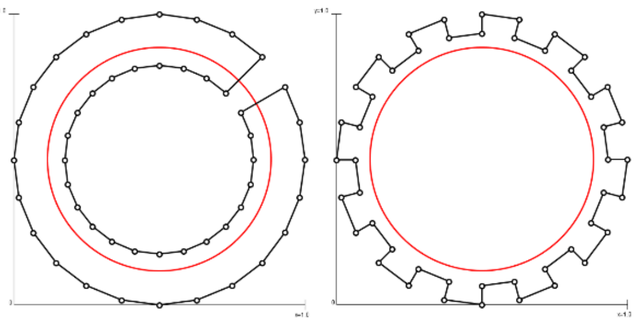
本論文では動的な環境を再現するために, 図 1 のような 2 重円の同心円状に同数の都市が均等に並んだ都市配置の TSP とした. この都市配置の内側の都市郡を内円都市, 外側の都市郡を外円都市と呼ぶ. この問題は内円都市と外円都市との半径比によって最適解が図 1 のような C 型, O 型のように変化する. 外円都市の半径を固定して内円都市の半径を変化させたとき, C 型と O 型の最適解が入れ替わる境界線の半径を R する. そのとき内円都市を R より内側に配置すると最適解は C 型になり, R より外側に配置すると最適解は O 型となる. このときのそれぞれの型の最短経路C_c, C_oは式(10), 式(11)で得られる.

$$C_c = 2\left\{(R_o + R_i)(n-1)\sin\frac{\pi}{n} + (R_o - R_i)\right\} \quad (10)$$

$$C_o = n(R_o + R_i)\sin\frac{\pi}{n} + n(R_o - R_i) \quad (11)$$

ここでR_oは外円都市の半径, R_iは内円都市の半径を, n は内円都市と外円都市に配置されている都市数を表している.

本実験では最初に内円都市を R より内側に配置する. このとき R と内円都市との差を r とする. その状態で探索を行い C 型の最適解を求める. 最適解を求めた後解の収束を行うためにそのまま 100 サイクル探索を続ける. その後, 内円都市を R より r だけ外側に再配置し環境の変化を起こす. その状態から探索を再開し O 型の最適解が得られるまでのサイクル数にて評価を 1 試行としてシミュレーションを行った. なお本実験では, 都市の再配置が行われてから 5000 サイクルを超えた場合, 探索が局所解に陥り最適解を発見できなかったものとして探索を終了した. またシミュレーションに用いた MMAS と AS-RE のパラメータは, フェロモンの重み$\alpha = 1.0$, 可視化の重み$\beta = 5.0$, エージェントが 1 サイクルで落とすフェロモンの総量$Q = 100.0$, フェロモンの蒸発係数$\rho = 0.8$, τ_{ik}の影響度$\mu = 3.0 \times \tau_{min}/(\tau_{max} \times n)$とした.



C型　　O型

図 1:2 重円都市配置の最適解と局所解

6.1 結果

r の初期値を 0.1 からシミュレーションを行い, 100 試行し, r を 0.001 刻みで減少させ, $r = 0.017$まで施行した結果を図 2 に示す. 縦軸は再配置から最適解を得るまでのサイクル数, 横軸は境界線半径 R と内円都市との半径の差 r を示している. 図 2 から$r = 0.081$までは環境の変化が大きいため MMAS と AS-RE ともに 1 サイクルで最適解を求めることが出来ている. しかしそれより r が小さくなると環境の変化が小さくなり問題が難しくなるために MMAS と AS-RE ともに最適解を見つけるまでのサイクル数が増えていることがわかる. しかし AS-RE は MMAS より常に少ないサイクル数で最適解を発見することが確認できた. このことより既存手法より非定常環境への適応を図れたのではないかと考える.

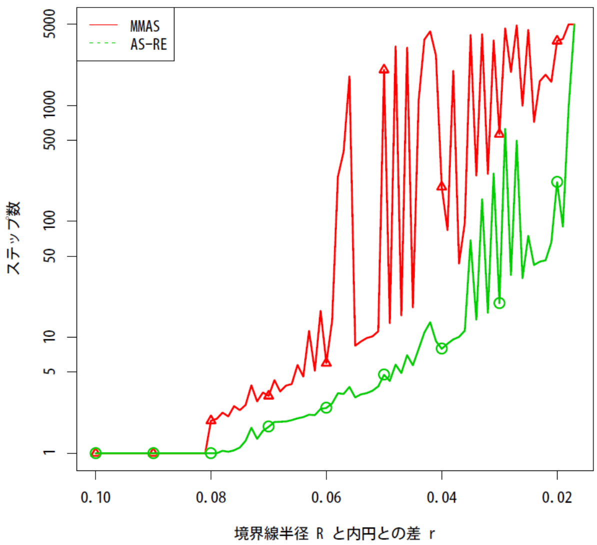


図 2:動的環境での実験結果

7. 結論

本研究では, 動的環境に適応するために相対評価を ACO に付加することで既存手法よりも性能の向上を図った. 結果として, 本実験で使用した動的環境において性能の向上を見ることが出来た. これは, 既存手法ではフェロモンの更新が絶対評価によって決定されていたものを, 提案手法ではエージェントが通った経路の変化量を通っていない経路にも影響させる, 相対評価によるものだと考える. この相対評価により解の収束が起こっていても解の再探索を促すことが出来たので既存手法より良い結果が得られたと考える. 今後の課題として, 他の動的環境における提案手法有効性と他のメタヒューリスティクスでの相対評価の有用性の検証をしたい.

参考文献

[M.Dorigo 96] M. Dorigo, V. Maniezzo, A. Coloni (1996): The Ant System:Optimization by a Colony of Cooperating Agents, IEEE Trans. SMC-Part B, Vol. 26, No. 1, pp. 29-41.

[山本 97] 山本芳嗣, 久保幹雄 (1997): 巡回セールスマン問題への招待, 朝倉書店.

[原 12] 原元司, 梶野大輔, 堀内匡 (2012): ACO による分割統治型 TSP 近似解法, 知能と情報（日本知能情報ファジィ学会誌）Vol. 24, No. 6, pp. 1101-1105.

[古川 12] 古川正志, 川上敬, 渡辺美知子, 木下正博, 山本雅人, 鈴木育男: メタヒューリスティクスとナチュラルコンピューティング, コロナ社,2012

[久保 09] 久保幹雄, J. P. ペドロソ (2009): メタヒューリスティクスの数理, 共立出版.

[古川 12] 古川正志, 川上敬, 渡辺美知子, 木下正博, 山本雅人, 鈴木育男 (2012): メタヒューリスティクスとナチュラルコンピューティング, コロナ社.

[Tversky 74]Tversky, A., Kahneman, D. (1974): Judgment under uncertainty: Heuristics and biases. Science 185 (4157), 1124–1131.