

アリの意思決定過程を考慮した Ant Colony Optimization による特徴選択

遠藤 博人[†] 穴田 一[†]

東京都市大学大学院 総合理工学研究科[†]

1. はじめに

近年、機械学習技術が発展しているが、学習データの中には学習に不必要なデータも含まれており、全データを用いて学習を行うと計算時間が膨大になり、精度も悪くなるという傾向がある。そのため、全データの中から学習に重要なデータ(特徴量)のみを選択する前処理(特徴選択)を行うことで、学習データの次元を削減する手法が提案されている。

特徴選択手法の1つに、Ant Colony Optimization (ACO)を適用した研究が存在する[1]。ACOは、実際のアリの採餌行動を参考にモデル化された手法である[2]。しかし、従来のACOには、重要なアリの意思決定過程が考慮されていないという問題がある。そこで本研究では、実際のアリが行う意思決定過程を導入したACOによる特徴選択手法を提案する。

2. 提案手法

本研究では、実際のアリが行う意思決定過程を導入したACOによる分類問題における特徴選択手法を提案する。

ACOは、グラフ上を複数のアリが探索することを繰り返して解を構築する手法であるため、特徴選択においても解空間を表すグラフを定義する必要がある。提案手法では、図1のように特徴選択を行うためのグラフを表現した。ここで、グラフのエッジ $C_{i,j}$ は特徴 F_i と特徴 F_j の相関係数の絶対値($0 \leq C_{i,j} \leq 1$)、 T_{F_i} は特徴 F_i とデータセットの説明変数であるクラスラベル(カテゴリカルデータ)との相関比($0 \leq T_{F_i} \leq 1$)である。各アリは、以下の式で示すFitness関数を最小化する特徴サブセットを発見するようにグラフ上を探索する。

$$Fitness = \frac{1}{mC_2} \sum_{i=1}^{m-1} \sum_{j=i+1}^m (1 - C_{i,j}) \times \frac{1}{m} \sum_{i=1}^m T_{F_i}$$

ここで、 m は指定された探索する特徴量の数であ

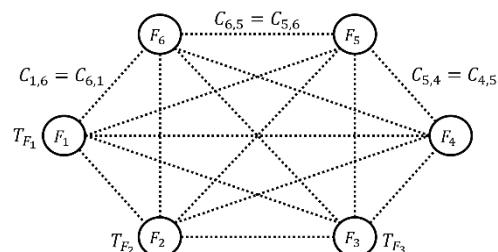


図1 特徴選択を行うためのグラフ表現

る。Fitness関数が最大となる特徴サブセットを発見することで、選択した特徴の冗長性の最小化と目的変数との関連性の最大化を実現できる。提案手法は、2.1節でフェロモンとMemoryの初期化し、2.2~2.4節を規定回数繰り返して探索を行い、2.5節で特徴サブセットを出力する。

2.1 フェロモンとMemoryの初期化

全エッジ(特徴間)のフェロモンを一律に定数 c で初期化する。また、各アリが解を記憶する領域であるMemoryをそれぞれ保存しており、各アリが1ステップ目に構築した解でアリ k のMemoryである $Memory_k$ を初期化する。Memoryは、2.3節の探索の際に活用する。

2.2 ローカルコミュニケーション

実際のアリは、探索の際に社会情報と個体情報に基づく意思決定を行い、多様な解探索を実現している[3]。社会情報は、フェロモンを用いたグローバルコミュニケーションや触覚タッチによるローカルコミュニケーションなど、他のアリから獲得した情報を指す。個体情報は、各アリが過去に通ってきた経路など、自身の採餌経験からの情報を指す。アリの意思決定過程は、多様な解探索を行う上で重要なアリの習性だと考えられるが、従来のACOでは考慮されていなかった。そのため、提案手法ではアリの意思決定過程をACOに導入した。この節では、アリの意思決定に必要なローカルコミュニケーションによって社会情報を獲得する。

まず、アリを各ノードに1匹ずつ配置し、ランダムにペアを組む。そして、ペア相手とローカルコミュニケーションを行い、相手のMemoryに記憶された解情報を獲得する。獲得した情報は探索の際に使用する。

2.3 探索

Feature Selection by Ant Colony Optimization Considering Ant Decision Making Process

[†]Hiroto Endo, Hajime Anada

Graduate School of Tokyo City University

各アリは、指定された特徴数を満たすまで探索を行う。 t ステップ目にアリ k が特徴 i にいる時、次に特徴 j を選択する確率 $P_{ij}^k(t)$ は以下の(1)式で表す。

$$P_{ij}^k(t) = \begin{cases} \frac{[\tau'_{ij}]^\alpha [\eta(F_j)]^\beta [T_{F_j}]^\gamma}{\sum_{l \in J_i^k} [\tau'_{il}]^\alpha [\eta(F_l)]^\beta [T_{F_l}]^\gamma} & \text{if } j \in J_i^k \\ 0 & \text{otherwise} \end{cases} \quad (1)$$

$$\tau'_{ij}(t) = \begin{cases} w_1 \cdot w_2 \cdot \tau_{ij}(t) & \text{if } F_j \in S_{PB}^k \text{ and } F_j \in S_{PB}^n \\ w_1 \cdot \tau_{ij}(t) & \text{if } F_j \in S_{PB}^k \\ w_2 \cdot \tau_{ij}(t) & \text{if } F_j \in S_{PB}^n \\ \tau_{ij}(t) & \text{otherwise} \end{cases}$$

$$\eta(F_j) = \frac{1}{s} \sum_{z \in F_j^k} (1 - C_{i,z}) \quad \text{if } F_j \in J_i^k$$

ここで、 $\eta(F_j)$ は選択済みの各特徴と特徴 j の相関係数を 1 から引いた値の平均、 T_{F_j} は特徴 F_j とクラスラベルの相関比、 J_i^k はアリ k の未選択特徴集合、 α, β, γ はそれぞれの情報の重み、 S_{PB}^k はアリ k の Memory に記憶された Personal Best の解、 S_{PB}^n は 2.2 節の会話相手 n の Memory に記憶された Personal Best の解、 w_1 ($w_1 > 1$) は S_{PB}^k の重み、 w_2 ($w_2 > 1$) は S_{PB}^n の重み、 $\tau_{ij}(t)$ は特徴 i, j 間のフェロモン量、 s は現在までに選択した特徴数、 F_j^k は現在までに選択した特徴集合である。まず、フェロモン情報 τ'_{ij} については、 S_{PB}^k と S_{PB}^n の一方もしくは両方に特徴 j が含まれていたら、特徴 j と繋がる全エッジ間のフェロモン量に重みを掛けて、探索の際に重視するようにした。以上のように、ローカルコミュニケーションで獲得した情報と探索を行うアリ自身の情報を探索に活用することで、実際のアリが行うような意思決定過程をモデルに表現している。

また、提案手法の $\eta(F_j)$ は、これまでに選択した全ての特徴と関連が低い特徴の値が大きくなるような式になっており、これまでに選択した全ての特徴との冗長性を最小化できる。

さらに、クラスラベルと特徴間の関係が考慮されていないと冗長性が低くても予測に関係しない特徴が選ばれてしまうという問題が生じる。提案手法では、 $\eta(F_j)$ の情報と併用して特徴 j とクラスラベルの相関比である T_{F_j} を探索に用いることで、選択した特徴の冗長性の最小化と目的変数との関連性の最大化を実現する特徴サブセットを選択できる。

アリ k は、(1)式に従って特徴を選択する度に、 $Memory_k$ に記憶された Personal Best を参照する。例えば、全 30 種類の特徴から 6 種類の特徴を選択する場合、アリ k の Memory である $Memory_k$ に

$\{F_2, F_5, F_{13}, F_{21}, F_{23}, F_{26}\}$ 記憶されているとする。アリ k が F_2 を出発し、次に(1)式に基づき特徴 F_8 を選択した場合、 F_8 は $Memory_k$ に存在しない新規の特徴であるため、 $Memory_k$ からスタートノードである特徴 F_2 を除いた特徴の集合である $\{F_5, F_{13}, F_{21}, F_{23}, F_{26}\}$ から F_8 との相関が低い上位 4 つの特徴を選択する。選択後の特徴サブセットの *Fitness* が元の $Memory_k$ の *Fitness* より小さければ探索を継続し、大きければその時点で探索を終了して、そのアリの解を新たに組み合わせた特徴サブセットとする。アリが序盤に発見した良い特徴サブセットを、記憶された Personal Best と組み合わせることで、有効活用することができる。

2.4 フェロモン更新

特徴 i, j 間のフェロモン $\tau_{ij}(t)$ は以下の式で更新する。

$$\tau_{ij}(t+1) = (1 - \rho)\tau_{ij}(t) + \frac{1}{n} \sum_{l=1}^n \delta\tau_{ij}^l$$

$$\delta\tau_{ij}^l = \begin{cases} Fitness_{S_l} & \text{if } F_i, F_j \in S_l \\ 0 & \text{otherwise} \end{cases}$$

ここで、 ρ はフェロモンの蒸発率、 n はアリの数、 S_l はアリ l の Personal Best、 $Fitness_{S_l}$ は S_l の *Fitness* である。各アリの Personal Best の中に、特徴 i, j が含まれていたら、特徴 i, j 間のフェロモンを更新する。ノード(特徴)ではなくエッジ(特徴間)にフェロモンを設定することで、特徴間の関係性の良し悪しを評価できる。

2.5 特徴サブセットの出力

フェロモン量が最大のエッジを結ぶノード(特徴)のどちらかをスタート地点として、そこからフェロモン量が最大の特徴を指定した数になるまで辿り、その特徴サブセットを出力する。

3. 評価実験

詳細な結果と考察は発表時に述べる。

参考文献

- [1] Tabakhi, S. and Moradi P. and Akhlaghian, F.. An unsupervised feature selection algorithm based on ant colony optimization. *Engineering Applications of Artificial Intelligence*, 2014, vol. 32, p. 112-123.
- [2] Dorigo, M. and Maniezzo, V. and Colnari, A.. *Ant System : optimization by a colony of cooperating agents*. *IEEE Transaction on SMC-Part*, 1996, vol. 26, no. 1, p. 29-41.
- [3] Kolay, S. and Boulay, R and Ettore, P.. *Regulation of Ant Foraging : A Review of the Role of Information Use and Personality*. *Frontiers in Psychology*, 2020, vol. 11, p. 1-7.