

無情報条件下での土壤汚染概況調査における試料採取地点の最適配置探索

京都大学大学院 学生会員 木内 智明
 京都大学大学院 正会員 米田 稔
 京都大学大学院 フェロー 森澤 眞輔
 三機工業株式会社 大塚 順基

1. はじめに

現在、土壤汚染の概況調査において、土壤のサンプリング密度は概ね 1000m²につき 1ヶ所で、試料採取は 1ヶ所につき、基本的には中心 1 地点及び周辺 4 方位の 5m から 10m までの間からそれぞれ 1 地点の合計 5 地点で行うという 5 地点混合方式¹⁾が提示されているが、その合理的な根拠については明らかにされていない。米田ら²⁾は、遺伝アルゴリズム(GA)³⁾を用いて 5 地点混合方式における試料採取地点の最適配置を求めているが、最適配置の探索に失敗するケースも見られた。それを踏まえて木内ら⁴⁾は、GAに加えてニューラルネットワーク(N.N.)⁵⁾を利用したアルゴリズムを用いて、より一般的な条件下において最適配置を探索した。本研究ではさらに、より性能の高いアルゴリズムを構築するとともに、汚染分布の空間的相関構造が未知であることも考慮した上で、任意の地形・任意の地点数において最適配置探索を行うことを試みる。

2. 評価関数の設定

2 次元座標を表すベクトルを X 、水平 2 次元空間で分布する確率変数を $Z(x)$ とする。 $Z(x)$ に統計的弱定常性と等方性を仮定すると、 $Z(x)$ の期待値 m は場所によって変化せず、 x_1 と x_2 での $Z(x)$ の値 $Z(x_1), Z(x_2)$ の共分散は、 x_1, x_2 間の距離 h のみの関数 $C(h)$ となる。 σ を $Z(x)$ の標準偏差とすると、相関係数 $r(h)$ は $r(h) = C(h)/\sigma^2$ で定義されるが、その関数形として本研究では、指数型相関関数形 $r(h) = \exp\left(-\frac{h}{L}\right)$ を適用した。ここで L は、相関の強さを決定するパラメータである。 Z_K を領域全体 S での空間平均、 Z_N を $Z(x_{N,j})$ ($j=1, 2, \dots, n$) の n 個の地点での算術平均とし、 Z_K を Z_N で推定するものとする、有限個の測定地点での平均値 Z_N と Z_K との差を 2 乗したものの期待値は米田らの研究の(13)式で表されるが、これを L の関数 $g(L)$ として表す。木内らは L の値を 3, 15, 60m とした場合についてそれぞれ $g(L)$ を評価関数として配置の優劣を評価したが、実際の概況調査では L の値は事前には不明である。いま、 L は正の値しかとらず、他には何ら L の確率分布に関する情報が与えられていないとすると、このような場合、ベイズ統計理論では、その対数値が一様分布するという無情報事前分布がよく採用される。本研究でも L の確率分布として、その対数値が一様分布すると仮定した。そのため本研究では、次式で表される $g(L)$ の期待値によって配置を評価することとした。

$$\int_{L_1}^{L_2} p(L) g(L) dL = \frac{c^2}{S^2} \int_S dS \left\{ \text{Ei} \left(-\frac{|X - X'|}{L_1} \right) - \text{Ei} \left(-\frac{|X - X'|}{L_2} \right) \right\} dS' + \frac{c}{n^2} \sum_{j=1}^n \left\{ \text{Ei} \left(-\frac{|X_{N,i} - X_{N,j}|}{L_1} \right) - \text{Ei} \left(-\frac{|X_{N,i} - X_{N,j}|}{L_2} \right) \right\} - \frac{2c}{nS} \sum_{i=1}^n \left\{ \text{Ei} \left(-\frac{|X - X_{N,i}|}{L_1} \right) - \text{Ei} \left(-\frac{|X - X_{N,i}|}{L_2} \right) \right\} dS \quad (1)$$

ここで $|x_{N,i} - x_{N,j}|$ は $x_{N,i}, x_{N,j}$ 間の距離、各積分は領域全体 S での面積分を表す。また $p(L)$ は L の確率密度関数、 c は $p(L) = \frac{c}{L}$ を満たす定数であり、 $\text{Ei}(x)$ は積分指数関数を表し、 L_1, L_2 はそれぞれ L のとりうる最小値と最大値である。 L についての積分区間の下限 L_1 は 0 をとり、上限 L_2 は、

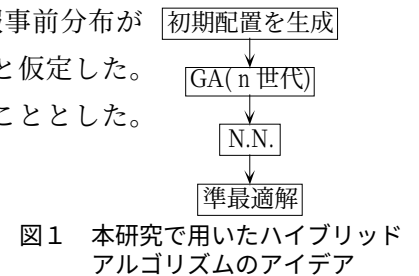


図1 本研究で用いたハイブリッドアルゴリズムのアイデア

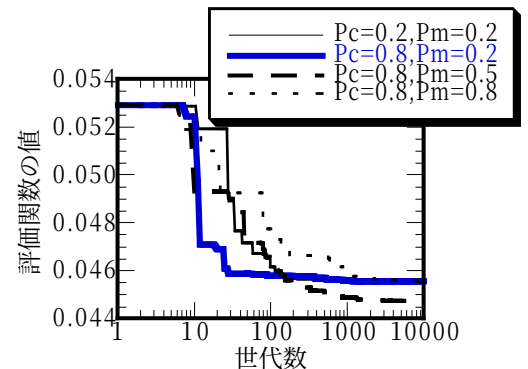


図2 GAの性能比較 (L=15m、個体数=1000)

キーワード：土壤汚染、試料採取地点、最適配置、遺伝アルゴリズム、無情報事前分布、空間的相関
 連絡先：京都市左京区吉田本町 (Tel. 075-753-5156 Fax. 075-753-5066)

式(1)の値の差が領域内の n 個の観測地点の位置に依存しないと見なせるように、指数型相関関数 $\exp\left(-\frac{h}{L}\right)=0.99$ を満たす L の値である $L_2=100h$ とした。 h としては対象領域における 2 地点の最大長さをとった。

3. 最適配置探索の手法

様々な形状の領域において、式(1)の値を最小にする n 個の試料採取地点配置の探索を、GA と N.N. を利用したアルゴリズムとを組み合わせたハイブリッドアルゴリズム(図1)により行った。GA 部分での解の選択方法としては米田らと同じくランク方式とエリート主義戦略を採用した。GA における重要な操作である交叉と突然変異とのそれぞれの発生確率 p_c と p_m の値の組み合わせによる評価関数の値の変化の例を図2に示す。図2では、 $(p_c, p_m)=(0.8, 0.5)$ のものが最も評価関数の値を下けているが、探索の初期(10~100世代くらい)においては $(p_c, p_m)=(0.8, 0.2)$ のものが評価関数の値をより早く減少させている。GA の「探索初期における大域的探索に優れている」という特長を踏まえ、局所的探索に優れる N.N. を利用したアルゴリズムとハイブリッド化させることを考えて、本研究では $(p_c, p_m)=(0.8, 0.2)$ を採用した。また、N.N. アルゴリズムにおいて見られる最急勾配法を利用したアルゴリズムは木内らにおいて構築されたもの(図3)を使用した。図3中の L_{ca} の値は、本研究では計算時間が短時間で済む点を重視して $L_{ca}=1$ を採用した。

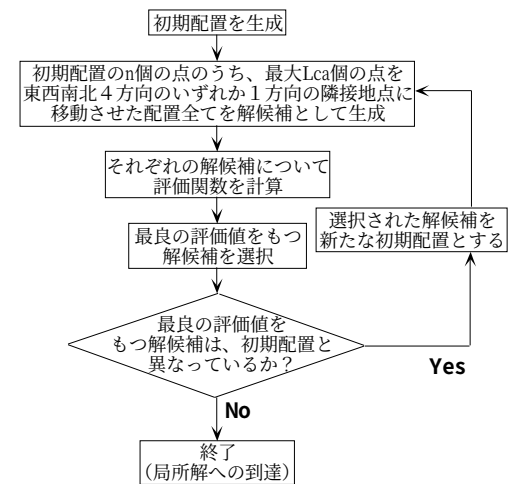


図3 本研究で用いたニューラルネットワークのフローチャート

4. 種々の形状の領域における試料採取地点の探索

数種類の形状の領域について最適配置の探索を行った結果を図4に示す。(1)は現行の5地点混合方式と同じく面積が1000m²の正方形領域における5地点の配置であるが、領域の中心に1点、残りの4点是对角線上にそれぞれ中心から約13m離れた配置となった。(2)は(1)の領域の縦方向の長さを2倍、横方向の長さを0.5倍にした領域における配置である。この形状の領域においては、(1)のような対角線上ではなく、領域の縦方向の中央線上に、領域の中心に1点、残りはそれぞれ12.5m~13m間隔で並んだ。(3)は任意形状の領域に適用した一例として、図のようなL字型の領域における5地点の配置である。さらに(4)は、(3)と同じ形状の領域において、7地点での配置の探索を試みた結果である。

5. まとめ

本研究では無情報下での土壤汚染概況調査における試料採取地点の最適配置を決定するために、有限個の観測値の平均値と領域全体での空間平均値との差の期待値を評価関数として、GAとN.N.を利用した最適化アルゴリズムを構築し、種々の形状の領域に対して最適配置探索を行った。今後さらにアルゴリズムの性能を向上させるとともに、サンプリングに関する経済性なども評価に組み込むことによって、より現実の土壤汚染調査に即したアプローチを目指していく。

参考文献

- 1) 環境庁水質保全局水質管理課土壤農薬課監修；土壤・地下水汚染対策ハンドブック, p.114, 公害研究対策センター, 1996
- 2) 米田稔、森澤眞輔、西村留美；「土壤汚染概況調査の5地点混合方式における試料採取地点最適配置」土木学会論文集、No.622/VII-II, pp.51-58、1999
- 3) (例えば) L. デービス；「遺伝アルゴリズムハンドブック」森北出版、1997
- 4) 木内智明、米田稔、大塚順基、森澤眞輔；「土壤汚染概況調査における試料採取地点の最適配置」土木学会第54回年次学術講演会講演概要集 第7部、pp.546-547、1999
- 5) (例えば) 武藤佳恭；「知識・情報・メディア／シリーズ ニューラルネットワーク」産業図書、1996

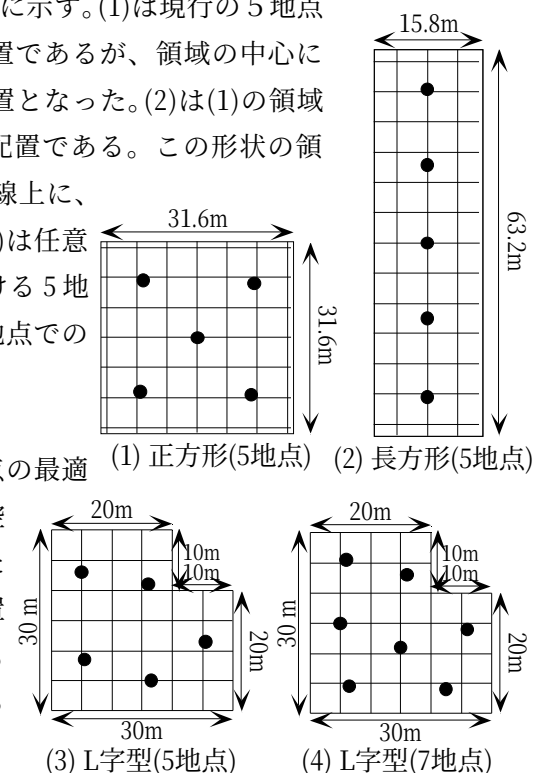


図4 最適配置探索の結果