

# 土地利用マイクロシミュレーションにおける 推定データの共生進化に基づく評価

Evaluation of Estimated Micro-Data in Land-Use Micro-Simulation using Symbiotic Evolution

大谷 紀子\*<sup>1</sup>    杉木 直\*<sup>2</sup>    宮本 和明\*<sup>1</sup>  
Noriko Otani    Nao Sugiki    Kazuaki Miyamoto

\*<sup>1</sup> 東京都市大学環境情報学部

Faculty of Environmental and Information Studies, Tokyo City University

\*<sup>2</sup> (株) ドーコン総合計画部

Research & Planning Department, Docon Co., Ltd.

Land-use micro-simulation models deal with detailed attributes of a household and its location which this study calls micro-data. Since an observed set of micro-data is usually not available, “synthetic population” is created from general accessible aggregate data provided by the national census with additional information obtained by sample survey. The validity of estimation method is evaluated by the goodness-of-fit between an estimated and the observed sets of micro-data. However, the goodness-of-fit cannot be calculated only with a conventional algorithm for an ordinary size of micro-data. In this paper, we propose a method based on symbiotic evolution to find out the best approximate value of the goodness-of-fit. Symbiotic evolution is a kind of GA characterized by parallel evolutions of both partial solutions and whole solutions. The effectiveness of the proposed method has been confirmed with the applications to the person trip survey data in Sapporo metropolitan area.

## 1. はじめに

土地利用マイクロシミュレーションモデルは、都市圏における土地利用と交通の詳細な変化の記述手法として、都市計画関連の各種分析への活用が進められている。住宅立地モデルのような世帯を対象としたマイクロシミュレーションの場合、各世帯の構成員数、各構成員の年齢、居住ゾーン、住宅タイプなど、複数の属性を定義した基準年マイクロデータをもとに、シミュレーションを実施する。しかし、個人や個別世帯に関するデータの入手は一般に困難であり、プライバシー保護の観点からも望ましくない。したがって、国勢調査などの入手可能な集計データと、追加的に個別世帯の情報を提供するサンプル調査結果から、個別世帯の各属性値を推定して基準年マイクロデータを作成する。以下では、実際のサンプル調査により得た実マイクロデータを観測データとして取り扱い、その観測データに対して推計作業を行なって得たマイクロデータを推定データとしてその推計精度の評価を行なう。

データ推定手法としては、IPF 法やモンテカルロサンプリングによる手法など複数の手法が提案されている [Pritchard 09, 杉木 09a, 杉木 09b, Miyamoto 09, Miyamoto 10]。マイクロシミュレーションモデルの有効性は基準年マイクロデータに依存するため、より実態に即したデータの推定が必要である。推定データの観測データに対する適合度を各データ間の乖離量距離和の最小値と定義し、異なるデータ推定手法の再現性能評価が試みられている。しかし、適合度算出には組合せ爆発の問題が伴うため、一般的な都市モデルにおけるデータ規模を想定した場合、厳密な適合度の算出は現実的に不可能である。

本稿では、大規模データにおける適合度の近似値を算出する方法として、共生進化に基づく手法を提案する。共生進化は遺伝的アルゴリズム (Genetic Algorithm, 以下 GA) の一手法であり、単純 GA よりも多様な解候補からの探索を実現できる点が特長である。処理対象であるマイクロデータの特徴を踏まえ、データ推定手法の比較評価に有用な適合度算出を目指す。

## 2. マイクロデータ

本節では、対象問題の定義、および提案手法で用いる世帯タイプについて説明する。

### 2.1 マイクロデータの属性と表現

マイクロデータの属性としては、前節で述べたように多様なものが考えられるが、本稿では、杉木ら [杉木 09b] に従い、世帯構成員の年齢、性別、世帯主との関係を属性とするマイクロデータを扱う。杉木らは、世帯構成員の性別と世帯主との関係を以下の 20 カテゴリで表現している。

本人 (男)	息子 1	息子 2	息子 3	その他 (男)1
息子の妻	父	兄弟	孫 (男)	その他 (男)2
本人 (女)	娘 1	娘 2	娘 3	その他 (女)1
妻	母	姉妹	孫 (女)	その他 (女)2

観測データ集合  $A$  と推定データ集合  $E_j$  は、それぞれ 1 世帯分のマイクロデータを要素として、式 (1) および式 (2) で表される。各要素は 20 次元ベクトルとなる。

$$A = \{a_i = (a_{i1}, a_{i2}, \dots, a_{i20}) \mid 1 \leq i \leq N\} \quad (1)$$

$$E_j = \{e_i^j = (e_{i1}^j, e_{i2}^j, \dots, e_{i20}^j) \mid 1 \leq i \leq N\} \quad (2)$$

ここで、 $N$  はデータ数、 $a_{ik}$  は観測データ集合における  $i$  番目の世帯の  $k$  番目のカテゴリに属する世帯構成員の年齢、 $e_{ik}^j$  は  $j$  番目の推定データ集合における  $i$  番目の世帯の  $k$  番目のカテゴリに属する世帯構成員の年齢を表す。当該カテゴリに属する世帯構成員が不在の場合、 $a_{ik}$  と  $e_{ik}^j$  の値は不在を表す数値 999 とする。 $i$  を世帯番号と呼ぶ。

### 2.2 推定データの適合度

観測データ  $a_i$  と推定データ  $e_k^j$  の距離  $Dis(a_i, e_k^j)$  は、式 (3) により算出される。計算時のオーバーフローと不在世帯構成員の過度の影響を回避するため、成分の差の二乗に上限値  $D_{max}$  を設けている。

$$Dis(a_i, e_k^j) = \sqrt{\sum_{l=1}^{20} \min\left((a_{il} - e_{kl}^j)^2, D_{max}\right)} \quad (3)$$

連絡先: 大谷紀子, 東京都市大学環境情報学部

〒 224-8551 横浜市都筑区牛久保西 3-3-1, 045-910-2938

E-mail: otani@tcu.ac.jp

表 1: 世帯タイプごとの世帯数

世帯タイプ	世帯数
男単身	2197
女単身	3225
夫婦	4996
夫婦 + 息子	1518
夫婦 + 娘	1582
夫婦 + 息子 2 人	668
夫婦 + 息子 + 娘	1311
夫婦 + 娘 2 人	587
その他 (世帯主男)	2201
その他 (世帯主女)	1109
合計	19394

推定データ集合  $E_j$  の適合度  $Fit(E_j)$  は式 (4) で定義される.

$$Fit(E_j) = \min_{\sigma \in S_n} \sum_{i=1}^N Dis(\mathbf{a}_i, \mathbf{e}_{\sigma(i)}^j) \quad (4)$$

ここで,  $S_n$  は集合  $\{1, 2, \dots, N\}$  から集合  $\{1, 2, \dots, N\}$  へのすべての全単射の集合を表し,  $\sigma(i)$  は全単射  $\sigma$  による  $i$  の像を表す. 式 (4) による適合度計算は,  $N!$  種類の全単射から, 距離和を最小とするような全単射を探索する問題といえる. すなわち, 距離和が最小となるように観測データ集合の各要素を推定データ集合のいずれかの要素と対応付ける組合せ最適化問題である. データ数を都市数, 適合度を総移動距離とする巡回セールスマン問題ともいえる.

### 2.3 世帯タイプ

世帯構成員の年齢, 性別, 世帯主との関係を式 (1), 式 (2) の 20 次元ベクトルで表すと, 各要素における 999 の存在状況により, 世帯を分類することができる. このときのカテゴリを世帯タイプと呼ぶ. 道央都市圏パーソントリップ調査で得られた 19394 世帯, 46500 人分のデータを分類したところ, 8 種の世帯タイプにおいて世帯数が 500 を超えた. 残りの世帯を世帯主の性別で分けて 2 種の世帯タイプと見なすと, 各世帯タイプの世帯数は表 1 のようになる.

観測データと推定データの距離は, 世帯タイプの相違に大きく依存する. 世帯タイプを考慮してデータの対応付けを考えることで, 単純な巡回セールスマン問題を解くための手法を適用するよりも有用な解が得られると予想される.

## 3. 共生進化に基づく適合度算出

共生進化に基づいて距離和が最小となるような観測データと推定データの対応付けを決定し, 適合度を算出する手法を提案する. 共生進化は, Moriarty らが提案した GA の 1 手法である [Moriarty 96]. 共生進化の特徴は, 部分解を個体とする集団と, 部分解の組合せを個体とする全体解集団を保持し, 両集団を並行して進化させる点にある. 部分解集団では解の部分的評価を行ない, 最適解に含まれ得る多様な部分解を生成する. 部分解のより良い組合せを全体解集団で学習することで, 1 集団を進化させる単純 GA よりも多様な解候補からの探索を行なうことができる. ILP や決定木生成への適用手法が提案されており, 有用性が確認されている [大谷 02, 大谷 04].

遺伝子型から表現型への変換時に, 世帯タイプの情報を使用する手法と使用しない手法の 2 種類を提案する. 後者はそのまま巡回セールスマン問題にも適用できる手法といえる.

### 3.1 部分解と全体解

提案手法では,  $A$  と  $E_j$  のデータの  $L_p$  組の対応付けを部分解集団の個体 (以下, 部分解個体) として表現する.  $A$  と  $E_j$  のデータの対応付けを世帯番号ペアと呼ぶ. 全体解集団の個体 (以下, 全体解個体) は  $L_w$  組の部分解個体の組合せとし, 1 個体で  $L_p \times L_w$  組の世帯番号ペアを表す.

部分解個体の染色体は長さ  $L_p \times 32$  のビット列とする. 16 ビットで 1 つの世帯番号, あるいは世帯番号を算出するための数値を表し, 図 1(1) に示すように, 1 個体で  $L_p$  組の世帯番号ペアを表現する. 全体解個体の遺伝子は部分解個体を参照するポイントであり, 染色体は長さ  $L_w = \lceil N/L_p \rceil$  のポイント列である. 全体解個体の例を図 1(2) に示す.

全体解個体から世帯番号ペアを生成するアルゴリズムを表 2 に示す. 本アルゴリズムにより致死遺伝子の生成を回避する. (1) では世帯タイプの情報を使わず,  $g_{17}^j \sim g_{32}^j$  を推定データの世帯番号そのものとみなして世帯番号ペアを生成する. (2) では, 表 1 に掲げた 10 種類の世帯タイプを使用し, 観測データと推定データの世帯タイプが同一になる確率を高めている. ここで  $g_{17}^j \sim g_{32}^j$  は, 世帯タイプが観測データと同じ推定データの中での通し番号を表していると考えられる.  $g_{17}^j \sim g_{32}^j$  が 0 以下の値を取る場合には, 観測データと異なる世帯タイプの推定データと組み合わせることとし, 世帯タイプの異なるデータ同士を組み合わせるも表現できるようにする.

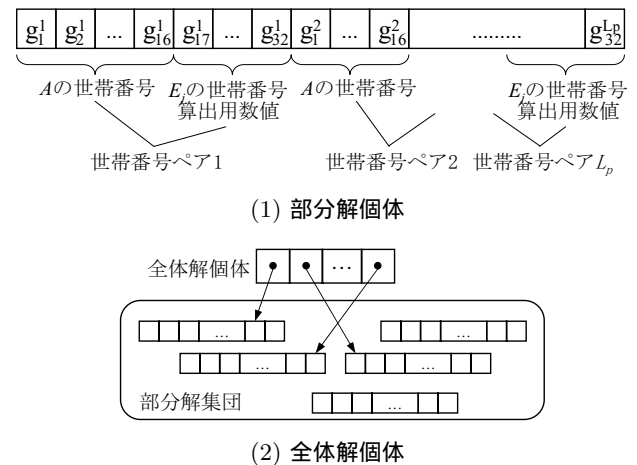


図 1: 両集団の個体

### 3.2 適応度と世代交代モデル

全体解個体の適応度は, 遺伝子が表す世帯番号ペアに基づいて算出される距離和とし, 適応度が小さいほど評価は高いものとする. 部分解個体は, 評価の高い全体解個体から参照されているほど評価が高いと判断する. 部分解個体  $I_p$  の適応度  $fit(I_p)$  の定義を式 (5) に示す.

$$fit(I_p) = \min_{I_w \rightarrow I_p} (fit(I_w) / avail(I_p, I_w)) \quad (5)$$

ここで,  $I_w \rightarrow I_p$  は全体解個体  $I_w$  が  $I_p$  を参照していること,  $fit(I_w)$  は  $I_w$  の適応度,  $avail(I_p, I_w)$  は  $I_w$  の世帯番号ペアの生成に使用された  $I_p$  の遺伝子の個数を表す.  $avail$  を加えることにより, 各遺伝子が全体解個体に与える影響の度合を反映させることができる.

部分解集団の世代交代では, [Moriarty 96] と同様にして, 上位半数の個体をそのまま次世代に残す. 下位半数の個体は, 上

表 2: 世帯番号ペア生成アルゴリズム

(1) 世帯タイプ不使用

```

for i:=1 to Lw {
  for j:=1 to Lp {
    idA:=i 番目の参照先の  $g_1^j \sim g_{16}^j$  が表す世帯番号;
    idE:=i 番目の参照先の  $g_{17}^j \sim g_{32}^j$  が表す世帯番号;
    if(idA も idE も世帯番号ペアで未使用)
      世帯番号ペア (idA, idE) を生成;
  }
}
for k:=1 to N {
  if(k は idA として世帯番号ペアで未使用) {
    idE:=min(idE として世帯番号ペアで未使用の数);
    世帯番号ペア (k, idE) を生成;
  }
}

```

(2) 世帯タイプ使用

```

for i:=1 to Lw {
  for j:=1 to Lp {
    idA:=i 番目の参照先の  $g_1^j \sim g_{16}^j$  が表す世帯番号;
    type:=fidA の世帯タイプ;
    n:=i 番目の参照先の  $g_{17}^j \sim g_{32}^j$  が表す数;
    if(n>0) {
      idE:=世帯タイプが type である推定データのうち、
        n 番目の推定データの世帯番号;
    } else {
      idE:=-n が表す世帯番号;
    }
    if(idA も idE も世帯番号ペアで未使用)
      世帯番号ペア (idA, idE) を生成;
  }
}
for k:=1 to N {
  if(k は idA として世帯番号ペアで未使用) {
    idE:=min(idE として世帯番号ペアで未使用の数);
    世帯番号ペア (k, idE) を生成;
  }
}

```

位四半数から選んだ 2 つの個体を親として 1 点交叉を行ない、生成された 2 つの子のいずれかと、2 つの親のいずれかで置き換える。すべての個体の遺伝子に対して確率  $p$  で突然変異を発生させ、次世代の個体とする。突然変異では当該遺伝子の値を反転する。

全体解集団の世代交代モデルとしては、[佐藤 97] で提案されている MGG (Minimal Generation Gap) モデルを採用する。MGG モデルは、局所解収束の回避と進化的停滞の抑制を意図して考案されたモデルである。集団からランダムに非復元抽出された 2 個体を親として 4 つの子を生成し、親と子の計 6 個体のうち、最良個体およびルーレット選択で選ばれた 1 個体の計 2 個体を次世代に残す。子の生成は 1 点交叉と突然変異に加え、遺伝子の順序入替えによるものも行う。表 2 のアルゴリズムによって遺伝子型を表現型に変換すると、遺伝子の順序が適応度を大きく左右するため、遺伝子の順序入替えによる子の生成も取り入れる。突然変異では、すべての個体の遺伝子に対して確率  $p$  で参照する部分解を変更する。

一世代の処理の流れを以下に示す。

1. 全体解集団の個体を評価
2. 部分解集団の個体を評価
3. 次世代の部分解集団を生成

表 3: パラメータ

パラメータ	値
突然変異確率 $p$	0.001
部分解集団の個体数	1000
全体解集団の個体数	1000
世代交代回数 $G$	5000

#### 4. 次世代の全体解集団を生成

初期集団を生成した後、上記の処理を  $G$  回繰り返す、全体解集団における最良個体の適応度を出力解とする。

### 4. 評価実験

提案手法の有効性を確認するため、道央都市圏パーソナルリッパ調査で得られたデータを用いて評価実験を行なった。なお、実験に用いたワークステーションのスペックは Intel Xeon 2.5GHz CPU、32GB RAM であり、パラメータの値は表 3 に示す通りである。

#### 4.1 適合度の正確さと計算速度

全探索によって正確な適合度算出が可能な小規模データ集合を用いて、提案手法の正確性と迅速性を検証する実験を行なった。13~17 世帯分のデータをランダムに抽出して、 $N = 13 \sim 17$  の観測データ集合と推定データ集合を 5 つずつ作成し、それぞれに関して共生進化に基づく 2 つの提案手法と全探索で適合度を算出した。部分解個体の染色体の長さ  $L_p$  は、データ数が小規模であることから 2 に設定した。

結果として、 $N = 13 \sim 17$  のすべてのデータ集合において、3 手法が同じ適合度を出力した。すなわち、共生進化に基づく手法で正しい適合度が算出できたといえる。全探索では、最小値をとる可能性がない場合には枝刈りを行なうようにしたが、計算時間は  $N = 13$  で 0.9 秒、 $N = 17$  で 1746 秒となり、 $N$  の増加とともに急激に増加した。一方、提案手法では、 $N = 13$  で 2.8 秒、 $N = 17$  で 3.5 秒となり、計算時間の増加は全探索と比較して微少であった。

#### 4.2 大規模データでの適合度

正確な適合度算出が困難な大規模データに関して得られる適合度を検証するための実験を行なった。4000 世帯分のデータを抽出して観測データ集合  $A$  とし、 $A$  の一部を加工して生成したデータ集合  $E_{a1} \sim E_{a5}$ 、 $E_{b1} \sim E_{b5}$  を推定データ集合とした。

データ集合  $E_{aj}$  は、以下のように  $A$  の世帯構成員の年齢を変更して生成した。

1.  $A$  からランダムに  $j \times 400$  世帯分のデータを選択する。
2. 選択したデータにおいて、年齢加工対象の世帯構成員をランダムに 1 名決定する。
3. 半数のデータにおいて、年齢加工対象世帯構成員の年齢から 5 を減ずる。
4. 残りの半数のデータにおいて、年齢加工対象世帯構成員の年齢に 5 を加える。

データ集合  $E_{bj}$  は、以下のように  $A$  の世帯構成員の性別を変更して生成した。

1.  $A$  からランダムに  $j \times 400$  世帯分のデータを選択する。
2. 選択したデータのうち単身世帯のデータについて、世帯主の性別を変更する。

表 4: 平均適合度

(1) 年齢変更データ

データ	世帯タイプ使用	世帯タイプ不使用	単純 GA
$E_{a1}$	634808.51	683285.02	799750.09
$E_{a2}$	699779.88	744279.42	901132.78
$E_{a3}$	762891.58	778486.92	916339.27
$E_{a4}$	797528.61	827519.81	959954.04
$E_{a5}$	817180.58	859172.53	984920.61

(2) 世帯構成員カテゴリ変更データ

データ	世帯タイプ使用	世帯タイプ不使用	単純 GA
$E_{b1}$	699729.17	726418.99	979246.19
$E_{b2}$	883027.10	920309.98	1026743.97
$E_{b3}$	994673.01	1033199.72	1142986.59
$E_{b4}$	1115525.26	1188136.00	1353977.38
$E_{b5}$	1187781.18	1269616.22	1413006.84

3. 選択したデータのうち非単身世帯のデータについて、年齢加工対象の世帯構成員を世帯主以外からランダムに1名決定し、世帯構成員不在の世帯主以外のカテゴリに変更する。

上記2つの手順で  $j = 1 \sim 5$  としてそれぞれ5つずつデータ集合を生成した。  $j$  が大きいほど観測データ集合からの変更度合いが高いデータ集合となる。

共生進化に基づく2つの提案手法、および単純 GA による適合度計算を各推定データ集合に関して10回ずつ繰り返したときの平均適合度を表4に示す。なお、部分解個体の染色体の長さ  $L_p$  は100に設定した。いずれの手法においても、観測データからの変更度合いが高くなるほど、高い適合度が算出されている。また、年齢変更データと世帯構成員カテゴリ変更データを比較すると、後者の方が高い適合度が算出されている。同カテゴリで年齢が異なるよりも、カテゴリが異なる方が距離は大きくなるので、妥当な適合度が出力されたと判断できる。

同データに対する適合度を比較すると、世帯タイプを使用した共生進化に基づく手法が最も低い適合度を算出しており、マイクロデータの特徴を生かした手法が有効であると考えられる。なお、計算時間は世帯タイプ使用の場合が約600秒、世帯タイプ不使用の場合が約450秒であった。

$E_{a1}$  の適合度を算出する際の2提案手法における最良全体解個体の適応度の推移を図2に示す。世帯タイプを使用する手法では、初期世代から世帯タイプを使用しない手法よりも低い適応度が得られており、進化を経てさらに低い適応度を獲得していることがわかる。

## 5. おわりに

土地利用マイクロシミュレーションにおける推定データの適合度算出手法を提案した。適合度算出は巡回セールスマン問題と同等の組合せ最適化問題であるが、マイクロデータの特徴を活用することで、より適切な値を算出することができた。今後は、さらに大規模なデータ集合への対応、および他の属性をも含むマイクロデータの適合度算出について検討する。

## 謝辞

本研究は、平成20~21年度科学研究費補助金基盤研究(B)「詳細属性情報を含む世帯の空間分布予測のためのマイクロシ

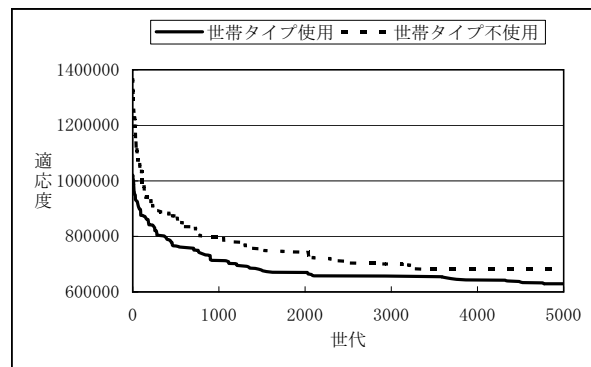


図 2: 最良全体解個体の適応度の推移

ミュレーションシステム」(課題番号 20360232) の助成を受けて実施した。ここに謝意を表したい。

## 参考文献

- [Miyamoto 09] Miyamoto, K. and Sugiki, N.: An Estimation Method of Household Micro-Data for the Base Year in Land-Use Micro Simulation, in *Proc. of CUPUM'2009*, 4.1-4 (2009)
- [Miyamoto 10] Miyamoto, K., Sugiki, N., Otani, N., and Vichiensan, V.: Agent-Based Estimation Method of Household Microdata for Base Year in Land Use Microsimulation, in *Compendium of the 89th Transportation Research Board Annual Meeting*, 10-3478 (2010)
- [Moriarty 96] Moriarty, D. and Miikkulainen, R.: Efficient Reinforcement Learning through Symbiotic Evolution, *Machine Learning*, Vol. 22, pp. 11-32 (1996)
- [大谷 02] 大谷 紀子, 大和田 勇人: 共生進化に基づく帰納論理プログラミングの予測精度の向上, *人工知能学会論文誌*, Vol. 17, No. 4, pp. 431-438 (2002)
- [大谷 04] 大谷 紀子, 志村 正道: 共生進化に基づく簡素な決定木の生成, *人工知能学会論文誌*, Vol. 19, No. 5, pp. 399-404 (2004)
- [Pritchard 09] Pritchard, D. and Miller, E.: Advances in Agent Population Synthesis and Application in an Integrated Land Use / Transportation Model, in *Compendium of the 88th Transportation Research Board Annual Meeting*, 09-1686 (2009)
- [佐藤 97] 佐藤 浩, 小野 功, 小林 重信: 遺伝的アルゴリズムにおける世代交代モデルの提案と評価, *人工知能学会誌*, Vol. 12, No. 5, pp. 734-744 (1997)
- [杉木 09a] 杉木 直, 宮本 和明, 大谷 紀子, Vichiensan, V.: 質的属性を含む初期マイクロ世帯データの推定手法, 第40回土木計画学研究発表会論文集, 18 (2009)
- [杉木 09b] 杉木 直, 宮本 和明, Vichiensan, V.: 土地利用マイクロシミュレーションにおける初期マイクロ世帯データ集合の推定手法, 第39回土木計画学研究発表会論文集, 135 (2009)