

## 進化する学習エージェント集団の動的環境への適応

山本 祐歌<sup>†</sup> 佐々木 貴宏<sup>††</sup> 所 真理雄<sup>†††</sup>

<sup>†††</sup> 慶應義塾大学大学院 理工学研究科計算機科学専攻  
<sup>†††</sup> ソニーコンピュータサイエンス研究所

本稿では、学習するニューラルネットワーク個体の集団が遺伝的アルゴリズムによって進化する過程を観察し、その際に個体の獲得形質の遺伝機構の異なる集団の動的環境への適応性について評価および議論する。そして獲得形質を遺伝しないダーウィン型の遺伝機構を持つ集団の方が、世代を通じて動的な環境に適応していくことが可能であり、変動する環境に対しても安定した挙動をみせること、更に未知の環境への潜在的な適応性も持つことを示す。

## Adaptation of Evolutionary Agents toward Changing Environments

Yuka Yamamoto,<sup>†</sup> Takahiro Sasaki<sup>††</sup> and Mario Tokoro<sup>†††</sup>

<sup>††</sup> Department of Computer Science, Faculty of Science and Technology, Keio University  
<sup>†††</sup> Sony Computer Science Laboratory

In this paper, we observe the adaptive process of the evolutionary agents in a simple abstract model, where neural networks capable of learning are evolved through genetic algorithms (GAs). Comparing two mechanisms of genetic inheritance, that are Darwinian and Lamarckian evolution, we show the following results. The population with no inheritance of acquired characters, i.e., Darwinian population, not only shows stable behavior against the oscillation of the environmental conditions and adaptability with respect to such changing environment, but also maintains a adaptational potential even toward completely new (unknown) environment.

### 1. はじめに

生物に代表される実世界のマルチエージェント系の設計あるいは動作原理は、人工物のトップダウン的に決定される設計とは全く異なるものであるが、実世界という予測不可能で複雑である動的な環境下において、極めて柔軟に適応している。従って開放分散環境などの動的な特性の強い環境で動作するエージェントシステムなどを構築する際に、これら実世界における系の持つメカニズムから学ぶ点が多い。実際、自然界のシステムに内在するダイナミクスをモデルとして、柔軟で適応的な情報処理の可能性を模索するようなアプローチが盛んに為されてきている [3]。本稿では学習するニューラルネットワークを個体とする集団を遺伝的アルゴリズムによって進化させる抽象的なモデルを用いて、集団が動的環境に適応していく過程について考察する。

自然界における生物の環境への適応は、「学習」と「進

化」という2つの過程として捉えることができる。生物個体の行動パターンは誕生時から固定されているわけではなく、その個体の生涯の間、経験を通じて生存に有利な行動や好ましいと思う行動をとる傾向が強くなる。個体の生涯というレベルでの適応、つまり「学習」が行われる。一方、生物個体は遺伝によって親の遺伝子の情報を受け継いで発生する。しかし、遺伝子の突然変異や交叉によって全ての子が親と全く同じ性質を持つわけではなく、他の個体に比べて少しでも有利に行動する個体を発生させる遺伝子は、自分の複製を次世代に残す可能性がより高くなる。こうした累積的な自然淘汰の過程で遺伝的な変異が積み重なることによって、世代をまたいだ集団レベルでの適応、つまり「進化」が起こる。適応は、これらの時間的、空間的に異なる2つの過程が相補的に作用していると考えられる。

ところで進化論の歴史では、遺伝現象と進化の原動力に関連してラマルキズムとダーウィニズムという二大思想が存在する。ラマルキズムは、進化の原動力を「獲得形質の遺伝」に置き、環境に対する応答として、学習な

<sup>†</sup> yuka@mt.cs.keio.ac.jp  
<sup>††</sup> sasaki@mt.cs.keio.ac.jp  
<sup>†††</sup> mario@csl.sony.co.jp

ど生物個体の生涯内に生じた適応的な変化が遺伝子に何らかの形で組み込まれ、それが子孫の変異に方向性を与えることで進化が推し進められるという概念である。それに対してダーウィニズムは、進化の原動力を「ランダムな変異に引き続く非ランダムな自然淘汰」とし、個体の変異自体には方向性がなく、たまたま適応的な変異を持って生じた個体が自然淘汰の過程で生き延びて繁栄するという考え方をとする。つまり、生物個体の生涯内に生じた適応的な変化の効果が遺伝子に組み込まれるという獲得形質の遺伝の考え方を否定する。今日の生物学や進化論においてはダーウィニズムが主流であり、ラマルキズムは異端もしくは誤りとされている。

しかし、工学的見地からはラマルク型の遺伝機構を考えることも可能であり [1][2]、獲得形質を継承するというのは直観的には有利であるように思われる。ところが、ダーウィン型とラマルク型の進化を変動する環境下でシミュレーションし比較した [5] では、ダーウィン型の遺伝機構を持つ集団は動的な環境に対しても安定した適応をする一方でラマルク型の集団は動的環境には対応できないことが示されている。本稿では、長期の一定期間ある環境下で進化してきた集団に突然新しい環境が与えられた際の適応の追従性と、各エージェントにあたるニューラルネットワークの出力値との関係からこれらの遺伝機構による違いを評価することで、[5]での議論をさらに発展させる。

## 2. 実験の設定

ここでは、本稿で用いる実験の設定について述べる。実験には佐々木ら [5] のモデルを用いた。以下に、論点を理解しやすくするために具体的なシナリオを想定した例題によって説明する。

100 個体のエージェントがそれぞれ初期値 500 の生命エネルギーを持って仮想世界に誕生する。ここで、各エージェントは環境から受け取る入力を元に行動を決定するフィードフォワード型のニューラルネットワークを持つとする (図 1)。このネットワークは各結合の強度を遺伝子として符号化した (つまり浮動小数点の配列として表現された) 染色体から直接エンコード法 [4][7] によって生成される。染色体上の各値は初期集団において  $\pm 0.30$  の範囲でランダムに初期化される。

仮想世界には、個体にとって有益な食糧と有害な毒物の二種類の物質が存在し、それぞれが特徴的な形状 (ビットパターン) を持つ (図 3)。物質が与えられるとエージェントはそれをニューラルネットワークに入力し、その出力値に応じてそれを「食べる」か「捨てる」かを決定する。ただし、この際に入力値をそのまま特定の行動に写像するのではなく、出力ニューロンの発火強度に応じて

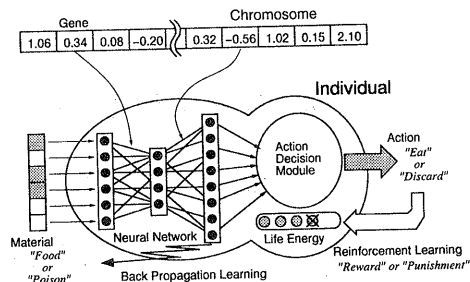


図 1 エージェントの構造

行動決定モジュールがボルツマン分布の下で確率的に最終的な行動を決定する。このような確率的な機構は、さらに有利な行動パターンを学習する可能性を残すために必要なものである。

食糧を食べるとエージェントの生命エネルギーは 10 単位増加し、そのパターンに対して「食べる」行動をとる確率が高くなるようにニューラルネットワークの学習が行われる。逆に、毒物を食べた場合は生命エネルギーが 10 単位減少し、そのパターンに対して「捨てる」行動をとる確率が高くなるような学習が行われる。食糧か毒物かに関わらず、エージェントが「捨てる」行動をとった場合には学習は行われない。このように、各エージェントは経験を通して自分の生命エネルギーを最大にするべく食糧と毒物の識別規則を学習する。エージェントの学習方法には、ニューラルネットワークの学習法の一つである誤差逆伝搬学習法に強化学習の枠組を被せたものを用いる。誤差逆伝搬学習法において、学習係数と慣性係数はそれぞれ  $\eta = 0.75$ ,  $\alpha = 0.8$  とした。

各個体エージェントは、食糧あるいは毒物を表すビットパターンを「生涯」の長さに対応する回数 (ここでは 400 回とした) 提示されて学習を行う。生涯を終えたエージェントは、生命エネルギーの残量を適応度として、それに応じた確率で選択され、ダーウィン型あるいはラマルク型の遺伝機構に基づいた遺伝的操作を施され、次世代に子孫を残す。ダーウィン型のエージェントの場合、子孫を作る際に親として使用される染色体は、生涯の間に修正された結合強度が全く反映されず、親から受け継いだ染色体がそのまま用いられる (図 2a)。一方、ラマルク型のエージェントでは、生涯の間に修正された結合強度をもとに再符号化された染色体が用いられる (図 2b)。遺伝的操作において、交叉点の数は 0 ~ 4、突然変異率は 5%、変異の幅は  $\pm 0.5$  からランダムにとるものとする。

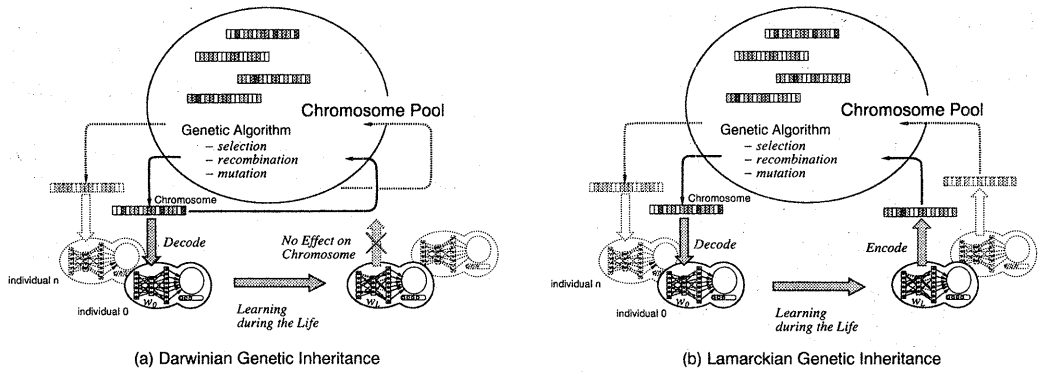


図2 ダーウィン型とラマルク型の遺伝機構

### 3. 実験の結果および検討

#### 3.1 実験1：静的な環境への適応性

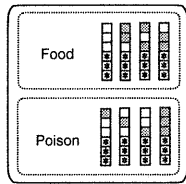


図3 実験1 — 環境の設定

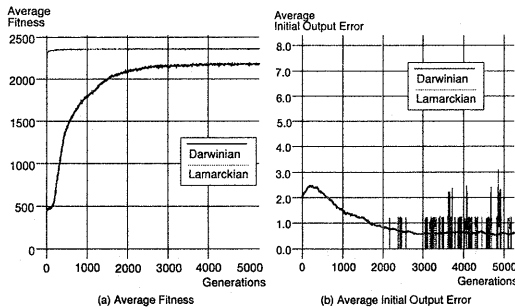


図4 実験1 — 平均適応度, 初期出力誤差平均の推移

はじめに、仮想世界における食糧と毒物を識別する規則が変化しない静的な環境を考える。この環境では図3に示すように、物質は6ビットのパターンからなる。“\*”はそのマスが白でも黒でもよいことを意味する。つまり、食糧と毒物とは上位3ビットで識別され、残りの3ビットは雑音ビットとなる。もちろん、各エージェントは雑音ビットが存在することさえも知らないものとす

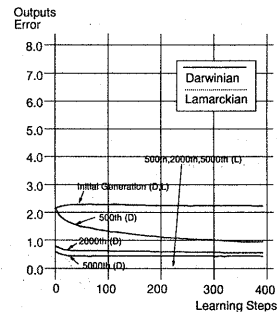


図5 実験1 — 世代経過に伴う学習特性曲線の変化

る。この環境下では、エージェントは雑音以外の3ビットを特定し、これら3ビットのパリティ問題を解くために必要な規則を学習することになる。

この環境においてシミュレーションを行い、エージェントを進化させたときの5000世代までの平均適応度の推移を図4(a)に、個体誕生時における初期出力誤差平均の推移を図4(b)に示した。出力誤差というのは、入力パターンに対するニューラルネットワークの理想的な出力値と実際の出力値との自乗誤差の和である。また、初期世代、500世代目、2000世代目、5000世代目のそれぞれの集団の学習特性曲線を図5に示した。これは、各世代のエージェントの一生間の学習の経過に伴う出力誤差の推移を表している。これらの図より、静的な環境ではラマルク型の遺伝機構を持つ方が圧倒的に学習効率が良いことがわかる。これはラマルク型の集団では、親の寿命によって途中で打ち切られた学習を子がそのまま継続することができるからであり、直観的な理解とも相違しない結果である。また、ダーウィン型の集団も、ラマルク型集団に比べて速度は遅いながらも、誕生時の初期出力誤差を減少し、すばやく環境を学習するように進化している。ここで、ダーウィン型集団の初期出力誤差が一度増加してから減少する点に留意していただきたい。

これについては後の実験の議論で触れることになる。

### 3.2 実験 2：動的な環境への適応性

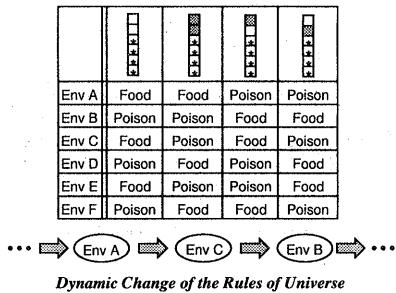


図 6 実験 2 — 環境の設定

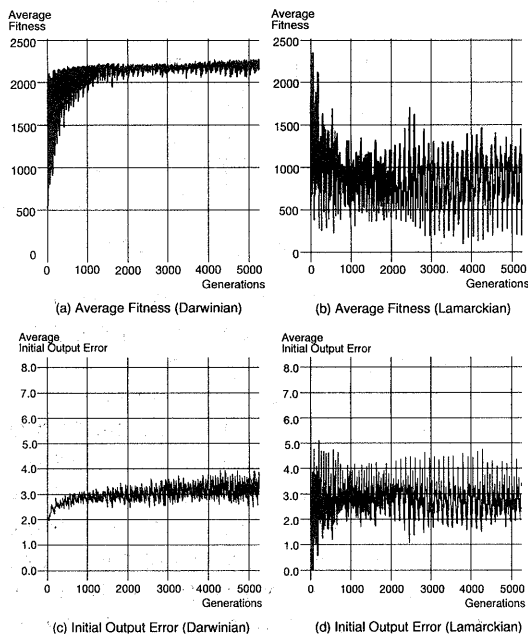


図 7 実験 2 — 平均適応度、初期出力誤差平均の推移

動的な環境に対する適応性を比較をするために、ここでは食糧と毒物を識別する規則自体が不規則に変化する環境を考える(図 6)。物質を示すビットパターンは 4 ビットの雑音を含む 6 ビットからなり、2 ビットの組み合わせによって食糧と毒物が識別される。以下では 20 世代毎に環境を変化させた時の結果を示す。

平均適応度と初期出力誤差平均の推移を図 7 に示す。これらより、ラマルク型集団の適応度は環境が変化する毎に大きく振動してしまい、うまく適応することができ

ないことが分かる(図 7b)。それに対してダーウィン型の集団では、振動はするものの世代を重ねるにつれて振幅が小さくなり、全体としての適応度は高くなる傾向を示している(図 7a)。

ここでダーウィン型の個体誕生時における初期出力誤差の世代に伴う推移(図 7c)を観察すると、静的環境で見られた(図 4b)ように減少するというよりも逆に増加する傾向にあり、敢えてある程度の初期出力誤差を保っているように見える。

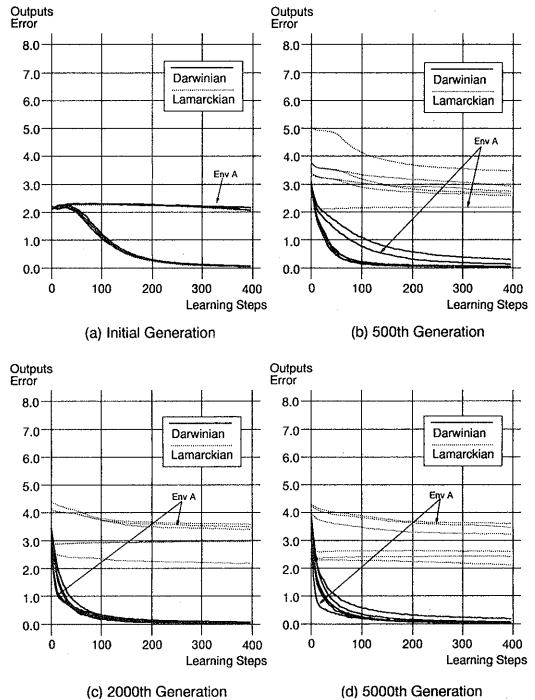


図 8 実験 2 — 世代経過に伴う学習特性曲線の変化

次に、初期世代、500 世代目、2000 世代目、5000 世代目の集団に Env A から Env F までのそれぞれの規則を与えたときの学習特性曲線を図 8 に示した。この図からも、ラマルク型はうまく学習ができないこと、ダーウィン型では初期出力誤差は世代を追うにつれて増加するが学習特性は向上していることがわかる。以上の結果から、ダーウィン型の集団は、動的な環境に対しては誕生時の出力誤差を大きくすることで様々な環境に対応し得る柔軟なネットワークを獲得していくと考えられる。この仮定を確かめるために、次の実験 3 を行う。

### 3.3 実験 3：未知の環境への適応性

ここでは、実験 1、実験 2 で進化させた各世代のエー

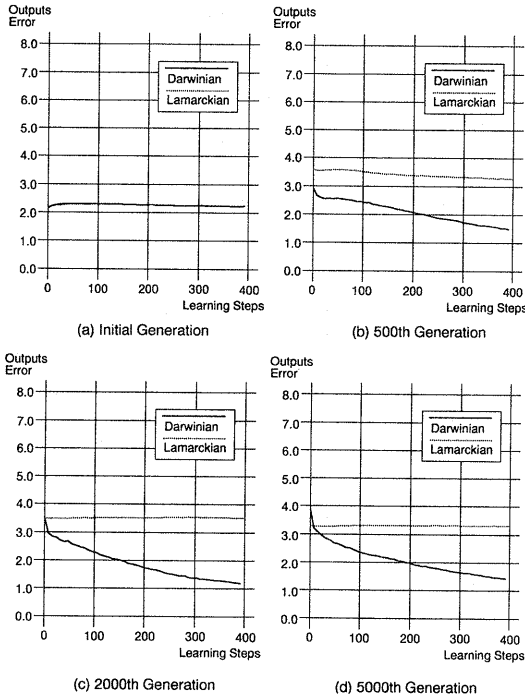


図9 実験3 — 世代経過に伴う学習特性曲線の変化(実験2の集団, 実験1の環境)

エージェント集団を用いて、それぞれに互いの環境を与えた際の学習特性を観察した。つまり、エージェントは自身が進化・適応してきた環境とは異なる環境(未知の環境)を突然与えられることになる。この実験によって未知なる環境に対する適応性を評価する。

まず、実験2の動的な環境で進化させた集団に実験1の環境を与えたものを図9に示す。ラマルク型の集団は、実験2で適応できなかったのと同様に、この新しく与えられた環境もまったく学習することができない。一方ダーウィン型の集団は、新しく与えられた環境であるにも関わらず、後の方の世代になるにつれ、ある程度の学習が行われることがわかる。

次に、実験1の環境で進化させた集団に実験2の環境を与えたものを図10に示す。図10で、ラマルク型の集団が適応できないということはこれまでの結果からも予想されることであるが、ダーウィン型の集団であっても静的な環境で長期間進化させたもの(2000世代目, 5000世代目)では、学習特性はかえって悪くなる傾向にある。静的な環境で進化したこのエージェントにとってはその環境が全世界であるため、その環境の規則にだけ適応すれば十分である。その結果、ある環境に特化するような個体に進化し、それ以外の環境への柔軟性が失わ

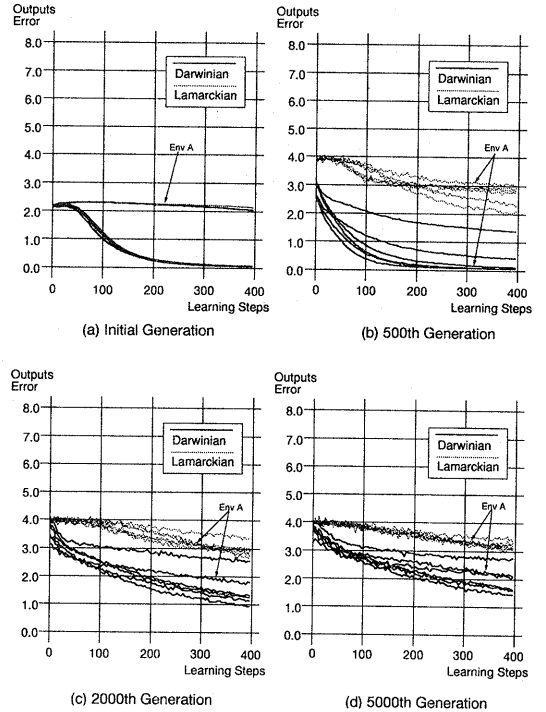


図10 実験3 — 世代経過に伴う学習特性曲線の変化(実験1の集団, 実験2の環境)

れるからだと考えられる。ここで注目すべきはダーウィン型の500世代目の集団の学習特性である。この世代のエージェントは、新たに不規則な識別規則を持つ環境を提示されてもかなり良い学習特性を示している。その一因として考えられるのが先にも述べた初期出力誤差に関係するものである。図4(b)を見ると、500世代辺りに初期出力誤差の一時的ピークがある。比較的優れた学習特性を示す世代と、このピークとが一致しているのである。

#### 4. 議 論

本稿で用いたモデルは学習、進化という現象の最小限の性質だけを抽出した単純なものではあるが、ここで得られた結果の中にも興味深い示唆をいくつか見出すことができると思われる。

実験の結果よりダーウィン型の遺伝機構を持つものは、まず初期出力誤差が増加し、その後静的な環境の場合は出力誤差は減少し、動的な環境の場合は高い値のまま維持されることが観察された。誕生時の出力誤差を増加させることはネットワークの柔軟性を保つことになり、それによってどのような環境にでもある程度適応できるように進化する。そして、環境が静的である場合にはその

環境に特化するようになり、初期出力誤差を減少させるようになる。一方、環境が動的である場合には初期出力誤差をそのまま高い値に維持することで、ネットワークの柔軟性を保ち、どのような環境にも適応できるような遺伝子を獲得することができている。このようにダーウィン型の集団は、どのような環境に対しても適応できるという頑健性、安定した適応性を持つということが示された。

ここで実世界を考えると、大部分の生物の遺伝機構はダーウィン型のものである。これは、生物が発生する時、DNA上の遺伝子型の情報からそれに対応する表現型が作られるが、逆にある表現型を作るための遺伝子型を構成するという、いわゆる“逆転写”というプロセスが非常に難しいためラマルク型の遺伝は不可能であるからというのが通説である。しかし、生物がダーウィン型の遺伝機構を進化の戦略としてとっているのは、ラマルク型の遺伝ができないという消極的な理由からではないかもしれない。つまり、動的な環境に対する頑健性があり、初期出力誤差を高い値に維持することで潜在的に「後天的に適応していくことのできる能力」を受け継いでいくとみられるダーウィン型の戦略の方が実世界の動的な環境で生存していくのに必要だったからと考えることもできる。

また、生物から着想を得て工学に応用しようとする人工的なシステムに対しても何らかの示唆を与え得ると思われる。開放分散環境などを考えると、環境は動的に変化するためそれに適応する能力が必要になってくる。そのために生物の機構を全く模倣してダーウィン型にしなければならぬという必要は無く、ラマルク型の親の獲得形質を継承することによる局所解への早い適応というのも魅力的である。人工的なシステムを構築するならばダーウィン型とラマルク型という両極端にこだわらず、両方の長所を取り入れた遺伝機構というのも考え得ること、その方向の探索も非常に興味深い [6]。

## 5. 結 論

本稿では、学習の効果と遺伝機構が進化に与える影響についてニューラルネットワークと遺伝的アルゴリズムを用いた抽象モデル上でシミュレーションを行い、評価した。得られた結果を以下に示す。

- ダーウィン型の遺伝機構の方が、ラマルク型の遺伝機構と比較して変動する環境に対して安定した挙動を示し、頑健性に優れる。
- ダーウィン型の遺伝機構は動的な環境に対して、誕生時の出力誤差を高い値に維持することで潜在的な適応性を保つ。

## 参 考 文 献

[1] John J. Grefenstette. Lamarckian Learning in

Multi-agent Environments. In *Proceedings of 4th International Conference on Genetic Algorithms and their applications (ICGA-91)*, pp. 303-310, 1991.

- [2] Akira Imada and Keijiro Araki. Lamarckian evolution of associative memory. In *Proceedings of 1996 IEEE The Third International Conference on Evolutionary Computation (ICEC-96)*, pp. 676-680, 1996.
- [3] Christopher G. Langton, editor. *Artificial Life: An Overview*. MIT Press, 1995.
- [4] David J. Montana and Lawrence Davis. Training Feedforward Neural Networks Using Genetic Algorithms. In *Proceedings of the 11th International Conference on Artificial Intelligence (IJCAI-89)*, pp. 762-767, 1989.
- [5] Takahiro Sasaki and Mario Tokoro. Adaptation toward changing environments: Why darwinian in nature? In *4th European Conference on Artificial Life (ECAL-97)*, pp. 145-153, 1997.
- [6] Takahiro Sasaki and Mario Tokoro. Adaptation under changing environments with various heredity rate of acquired characters: Comparison between darwinian and lamarckian evolution. In *2nd Asia-Pacific Conference on Simulated Evolution And Learning (SEAL-98)*, 1998. to appear.
- [7] Darrell Whitley and Thomas Hanson. Optimizing Neural Networks Using Faster, More Accurate Genetic Search. In *Proceedings of 3rd International Conference on Genetic Algorithms and their applications (ICGA-89)*, pp. 391-396, 1989.