

普遍文法の進化的ニューラルモデル

齊木規広[†] 松田聖^{††}

言語や文法の能力の獲得は人間特有の知的活動でありこれまでニューラルネットワークを用いたいくつかのモデルが提案されてきた。しかし、これらの能力はその全てを生後に一から学習した訳ではなく、人間の長い進化の中で獲得したものに負う部分も大きいと考えられる。そのようなものとして普遍文法が提案されている。普遍文法は人間が進化する中で獲得した脳の中に生得的に埋め込まれた、特定の言語に依存しない包括的な文法の知識であると考えられている。本論文では、ニューロ進化手法を用いることで、進化的ニューラルネットワークモデルを呈示し、普遍文法が進化の中で獲得されていく様子をシミュレーションで検証する。

Evolutionary Neural Network Model of Universal Grammar

Motohiro Saiki[†] and Satoshi MATSUDA^{††}

Abstract---Acquisition and performance of languages or grammar are the typical intellectual activities of human beings, and various models of these processes using neural networks have been proposed. These activities are assumed not only to be learned completely anew in each individual, but also to have been acquired over the long evolutionary history of human beings. The universal grammar is considered to be a comprehensive knowledge of grammar that has been acquired and hardwired in the brain during human evolution. By employing neuroevolution, we illustrate how the universal grammar might have evolved in the neural network using a genetic algorithm.

1. はじめに

人は生まれてから僅かな年数で言葉を上手く話せるようになる。言語の習得には環境によって変化する。例えば、日本で生まれ、アメリカで育つ環境では英語を習得すると思われるが、実際にはそうではない。上記の環境でも、日本語しか使わない場合は日本語を習得する。アメリカで生まれ日本語しか使わない環境であれば日本語を習得する。人が生まれてから初めて習得した言語を第一言語という。母国語を習得するという訳ではなく、環境次第で変化するため母語を習得することである。全ての人は、生まれてからの僅かな年数の内に文法的なことを勉強した記憶はないだろう。それにも関わらず私たちは特別意識せずに、他者とのコミュニケーションを取る為に使っている。言語の習得の研究をしている人達にも言語獲得の理由が多数あり、ある人は、僅かな年数の内に、耳から入る情報の中で文法的なことを学習していると言う。またある人は、生まれながらにして、言語を覚える際に特別な機関、システムを用いて学習をしていると言う。このことを普遍文法 (Universal grammar) と呼び、普遍文法があるため、言語を獲得できるとしている。前者は生まれてからの経験で言語を覚えるという理論であり、この理論に沿った研究で、ニューラルネットワークを用いた言語処理のモデルをエルマンが考案した。エルマンが考案した言語処理に対してのエルマンネットワーク (脳を模擬したニューラルモデルの一種) は、言語処理のシミュレーションを行うことができる[1]。本研究では、生まれながらにして、普遍文法を持つていたとの仮説に基づいているので、このままでは進化的な要素が抜けており、人間の進化の中で生得的に獲得したことへのアプローチができない。そこで、エルマンネットワークに進化的要素を加えるため、遺伝的アルゴリズムを用いることによって、普遍文法の進化的モデルを提案し、シミュレーションによってその妥当性を示す。

2. 普遍文法について

普遍文法とは、チョムスキーが提唱した言語学の一つの理論である、生成文法の重要なモノ、言語機能の初期状態についての理論のことである[2],[3]。生成文法では、言語とは人の外側にあるものではなく、人間の心、脳の中に存在しているとして、誕生時に言語に対して学習する機構、システム的なものがあるとしており、これにより人は言語を習得できるとしている。言語を習得するためには、ある程度の入力 (両親の言葉や周りで話されている言葉を含める) が必要で、その入力された言葉から普遍文法的作用により言語を習得している。そもそも普遍文法の始まりは、第一言語獲得

[†] 日本大学大学院生産工学研究科数理情報工学専攻
Graduate Course of Mathematical Information Engineering, Graduate School of Industrial Technology, Nihon University

^{††} 日本大学生産工学部数理情報工学科
Department of Mathematical Information Engineering, College of Industrial Technology, Nihon University

の際に起こる問題のアプローチである。その問題とは、刺激の貧困である。生まれたばかりの幼い子供が、周りから聞こえる言葉を入力として、言語を習得しようとしている。しかし、ここで問題がある。それは周りから聞こえてくるものは何も文法通り正しい文章だけではない。例えば、今、二人の人が会話をしている場合を考える。Aさんが「昨日ハンバーグ食べたんだけど、おいしかったさー」と喋ったことに対して、Bさんが「お店のハンバーグ？」と返したとする。ここで、Aさんの喋った言葉は、会話上普通に聞こえるが、文法として見てみると、主語が省略された文章になっている。Bさんも動詞が省略されている。日本語特有のことで主語を省略することや、動詞を省略することは言語を習得している人からすると、理解できることである。しかし、私たちは生まれてからきちんとした文法を学ぶことなく言語を習得している。幼い子供に話しかける時、きちんとした文法通りに喋りかける人はそうそういないだろう。それにも関わらず子供はきちんと言語を習得することができる。さらに刺激の貧困は文法に関することだけではない。それは正否判断の点である。耳から聞こえてくる言葉はただでさえちゃんとした文法的ではないので、正しい文法なのか、そうでないかの判断をしなければ学習にはならない。生まれたばかりの子供は、このような不完全な言葉が入力された時、不思議と正否判断ができていく。これは子供の喋ることを研究していると分かったことなのであり、子供は言葉足らずな喋り方をすると、文法的におかしい喋り方は絶対にしないことがわかった。これは子供が正否判断をして、文法を学習しているということになる。その理由として、大人が子供の喋ったことに対して間違いを指摘したりしないため、正否判断は子供が行っていることになる。このように、私たちは文法を勉強していない状態で成長する。子供が喋るようになるまでの短期間に聞いて育つ言葉は、不完全なものばかりで、文としても短いものである。それに加え、正否判断を教えてもらっている訳でないにも関わらず、子供は言葉が話すようになる。機械学習理論によれば、負例（正しくないことが明示された例題）が与えられない帰納的学習には限界があることが知られている[4]。以上のことから、言語を獲得しようとしている子供の脳の中にそれを可能にさせているなんらかのシステムが生得的に存在しているものと考えられる。それが普遍文法である。

3. 遺伝的アルゴリズム

3.1 遺伝的アルゴリズムとは

遺伝的アルゴリズム (GA) とは、ある目的に則した遺伝子を進化的に獲得する手法で、染色体の評価値により、操作を行うアルゴリズムである。その操作とは選択、交叉、突然変異、3つである。

3.2 選択

選択とは生物の自然淘汰をモデルとしている。染色体の適応度に応じて、ルーレッ

ト、ランク、トーナメント選択などがあり、選択によって適応度の最大値が下がらないように、次の世代に適応度が高い遺伝子を保存するエリート戦略などがある。

3.3 交叉

交叉とは、遺伝子群の遺伝子の列を交換することである。母集団の中から2つ染色体を選択し、交叉させることで、新しい染色体が生まれる。これを繰り返すことで、染色体を進化させていく。手法として一点交叉、二点交叉、多点交叉、一様交叉などがある。

3.4 突然変異

突然変異とは遺伝子のある箇所の値をランダムに変化させることで、局所解に陥るのを防ぐ効果がある。一般的に突然変異の確率は0.1~1%とされている。この確率が小さすぎると局所解に陥りやすく、逆に高すぎるとランダム探索になってしまう。

3.5 実数値 GA [5]

遺伝的アルゴリズムは、本来遺伝子の値が0と1で表現されていた。二進数で表現されているので、交叉によって新たな値になる可能性があった。例えば 00101、という値を持つ遺伝子があったとする。この遺伝子の一つでも変われば表現される値が変わる。交叉による操作でも、値が変わることから、探索空間を広げることができる。遺伝子の値を0と1ではなく実数を用いる GA があり、これを実数値 GA と呼ぶ。実数値 GA は遺伝子の値に実数を用いる。実数値 GA の手法はいくつかあるが、今回用いているのは染色体の遺伝子の値を-0.1 から 0.1 の範囲で乱数を取り、遺伝子の値を加減する手法である。遺伝子の値を僅かに加減することで少しずつ探索空間を広げていくことができる。

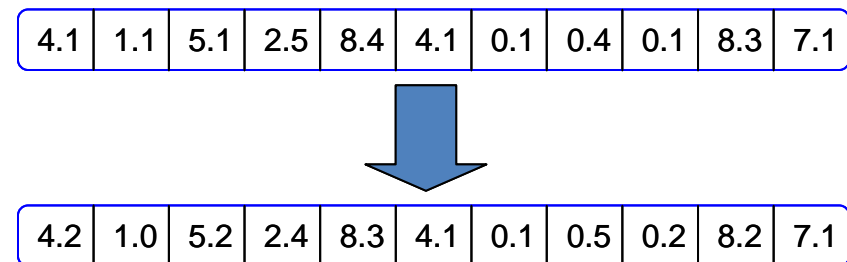


図1 実数値 GA

Fig.1 Real-coded Genetic Algorithms.

4. エルマンネットワーク[6]

以下の図はエルマンネットワークの図である。

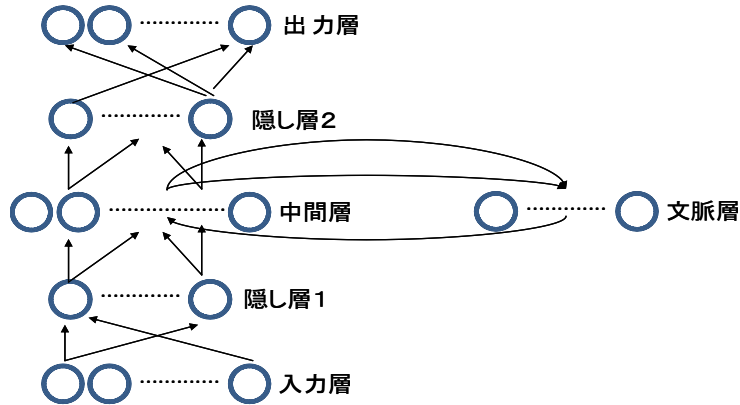


図2 エルマンネットワーク
Fig.2 Elman network.

エルマンネットワークとは、エルマンが考案したニューラルネットワークのモデルである。エルマンネットワークは、言語処理のモデルとして考案された。普通のニューラルネットワークとの違いは、文脈層による効果である。その効果とは、普通のニューラルネットワークでは、時系列が絡んでいる処理ができない。例えば、A, B, Xといった記号が、順番に入力層のニューロンが発火したとしても、Aに対して、Bに対して、Xに対してといった、個別の入力に対しての正しい出力をするようにしか学習しない。A, B, Xの一連の入力がされて始めて意味を成す入力に対しての学習を、文脈層を入れることにより実現している。この性質を用いると、自然言語処理に使える。例えば、日本語の文章で、「私は、明日学校へ行く」という文章を入力文章とすると、「私」「は」「,」「明日」「学校」「へ」「行く」という単語群に区切り、一つひとつを入力層と出力層のニューロンに対応づける。単語に対応しているニューロンが発火すると、その次の単語を出力することが正しい出力になる。例えば、文脈層には「私」が入力の時の情報が入っているとすると、次の単語の入力時に用いられる。次に「は」という単語に対応する入力層のニューロンが発火した時、前の入力であった「私」の情報と合わせ、「私は」という情報をもっていることになる。こうして文脈層に前の入力を保持し、文章を理解していくことになる。エルマンネットワークの学習方法はバックプロパゲーション（誤差逆伝播法）を使い、出力に対して教師信号から信号を受け、成

否を判断し、正しい出力になるように下の層へと誤差を修正していく学習方法を用いている。

5. 普通文法のニューラルモデル

以下の図は本シミュレーションで用いるフローチャートと日本語、英語の例文の図である。

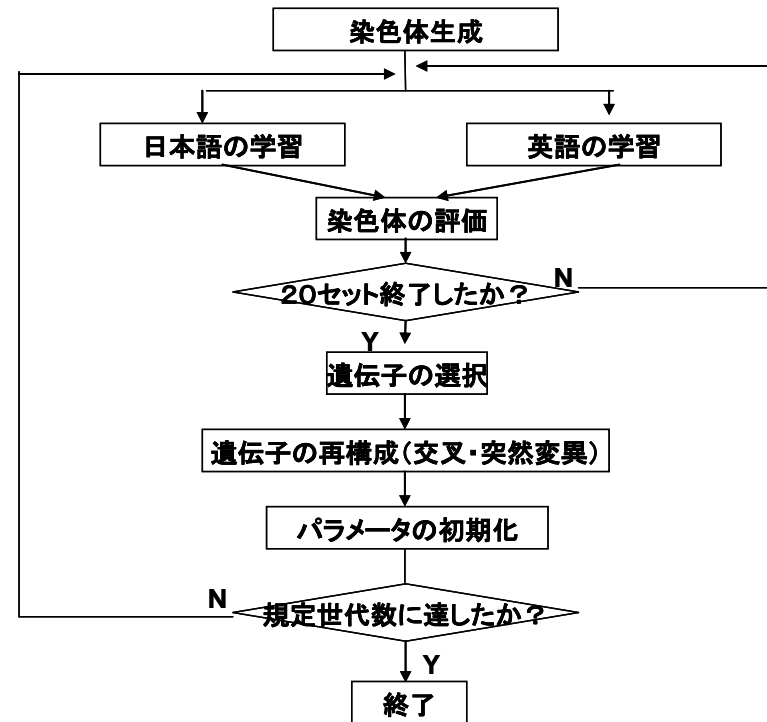


図3 フローチャート
Fig.3 flow chart.

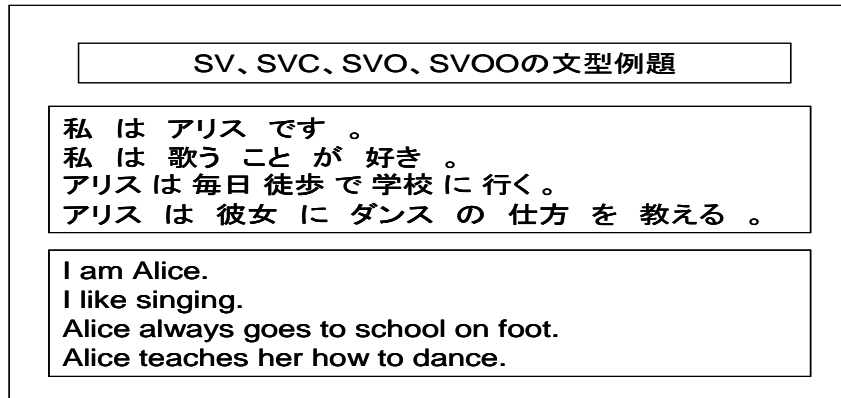


図4 日本語、英語の例文

Fig.4 Example of Japanese and English.

5.1 普遍文法のニューラルモデル

提案するモデルはエルマンネットと遺伝的アルゴリズムを用いるものである。使用するエルマンネットは入力層と出力層がそれぞれ26個、隠し層1、2は各11個、中間層と文脈層が各71個のニューロンから構成され、エルマンが行ったようにBPで学習し、各言語の知識（文法）が結合の重みとして形成される。これは個体の後天的な学習に相当する。又、人類の進化による普遍文法もやはりエルマンネットの結合の重みとして形成される。すなわち、普遍文法は後天的な学習と遺伝的な進化の相互作用によって、結合の重みとして形成されると考えられる。そこで染色体はエルマンネットの隠れ層1から中間層、中間層から隠れ層2への結合の重みから構成されるものとする。その理由は、これらネットワークの奥深くの結合の重みには個々の言語に対し独立の普遍文法の知識が形成されると考えられるからである。又、入力層から隠し層1、隠し層2から出力層の重みは個々の言語の知識を奥深くの普遍文法に変換する役割をもち、言語依存であると考えられる。

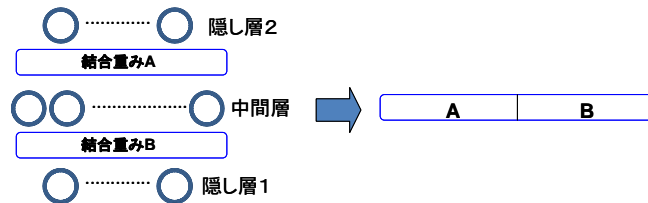


図5 遺伝的アルゴリズム

Fig.5 Genetic Algorithm

5.2 初期染色体の生成

結合荷重の内部パラメータは-1~1の範囲の中で乱数を取っており、日本語、英語ともに初期値のパラメータを保存する。その理由として、初期パラメータは乱数で取っているため、初期値が毎回変わってしまうのは初期値からの進化の変異を捕らえることができなくなるためである。初期値を再び日本語、英語のエルマンネットワークのシミュレーションを行う際に適用することで、初期値の乱数による違いがなくなり、結果の比較を容易にしている。

5.3 日本語及び英語の学習

入力層と出力層は単語一つひとつに対してニューロンに対応している。入力層の1番目が「私」、6番目が「アリス」など、ニューロンと対応付けられているので、例えば、入力層の1番目が発火状態だと、「アリス」の単語が入力されているということになる。学習の進み方は文章どおりに単語を次々に発火させ入力させていく（単語に対応するニューロンの値を1にする）。「私」が発火した次の入力は「は」となり、最後の「。」が入力されると、次の文章の先頭の単語へと入力される。バックプロパゲーションによる学習のタイミングは各単語が入力され、次にくる単語の予測を出力した時に行われる。シグモイド関数を用い、学習係数は0.1固定で、学習回数が増えるごとに減衰はさせていない。学習は全ての文章の単語を入力した時に、次の単語が正しく出力されるまで行われる。全て正しく出力されない場合は最初の文章の1単語目から繰り返し、4つの文章を完全に出力できるまで学習を続ける。ここで注意すべき点は1~4の文章を入力するが、1ループの間で2、3、4の文章を完璧に出力することができ、次のステップの時に1が完璧に出力したとしても、そこで学習は終了することではなく、1~4の文章を順番に入力していき、全ての文章に対して正しい出力をすることがシミュレーション終了条件となる。ここまですべて言語を学習することであり、エルマンが行ったことと同様である。

5.4 染色体の評価

染色体を遺伝子の初期値を-1~1の範囲から乱数で20セット生成する。20セット生成した結合荷重を一つずつ、日本語、英語のエルマンネットに適用する。各ネットワークでBPによる学習が終了した後に、評価値を計算する。評価値の優劣は、日本語、英語それぞれのシミュレーションにおけるバックプロパゲーションによる学習回数が少ない程良いとする。20セット分の結合荷重を適用したシミュレーションが終了するまで続ける。日本語、英語ともにバックプロパゲーションの学習回数は20万回を上限として、上限に達するとそこで強制的に終了させる。

5.5 染色体の選択

選択はエリート戦略とランク戦略を用いており、交叉は一点交叉、突然変異の確率は0.5%としている。実数値GAとしては、遺伝子群の値に対して、-0.1~0.1の値を加算している。この操作により探索範囲を初期値から少しずつ広げることができる。

具体的な操作は、日本語、英語のシミュレーションを行った後、評価値を算出し、母集団にエリート戦略を用い、評価値の高い染色体を2つ次世代に残し、残りの18個の染色体に対しランク選択を用いる。

5.6 染色体の再構成 (交叉, 突然変異)

選択された染色体に対して、一点交叉を行い、次世代の染色体を18個作り出す。その後実数値 GA により、遺伝子の値を-0.1~0.1の範囲の乱数を加減する。次に0.5%の確率で突然変異を行い、次世代の染色体を作り出す。

5.7 パラメータの初期化

次世代の染色体を作り出すと、パラメータを保存していた初期値に戻し、次世代の染色体を日本語、英語用のエルマンネットワークに適用し、規定数の次世代数になるまで上記の操作を繰り返し続ける。

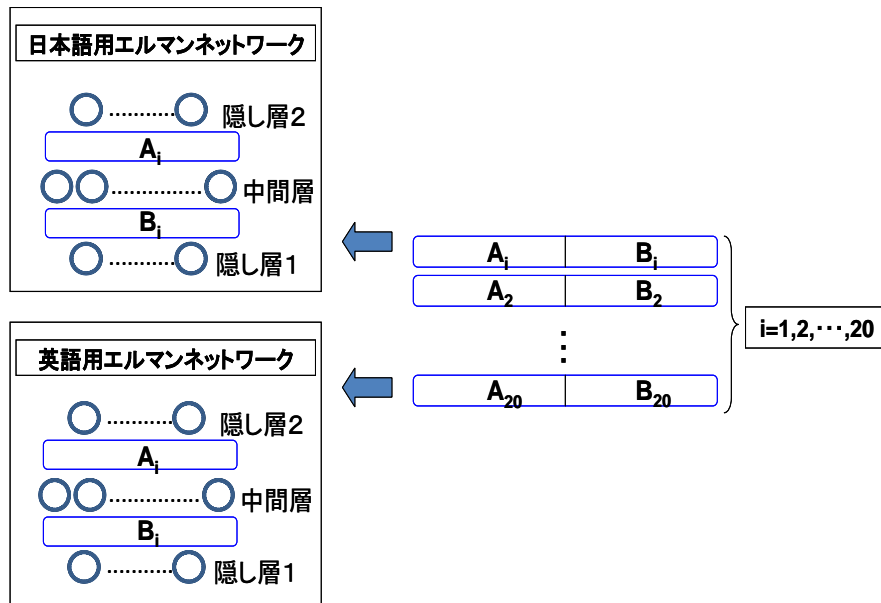


図6 染色体からエルマンネットワークの生成
Fig.6 Elman network produced by chromosome

6. 実験結果

今回の実験の結果を以下の図で示す。

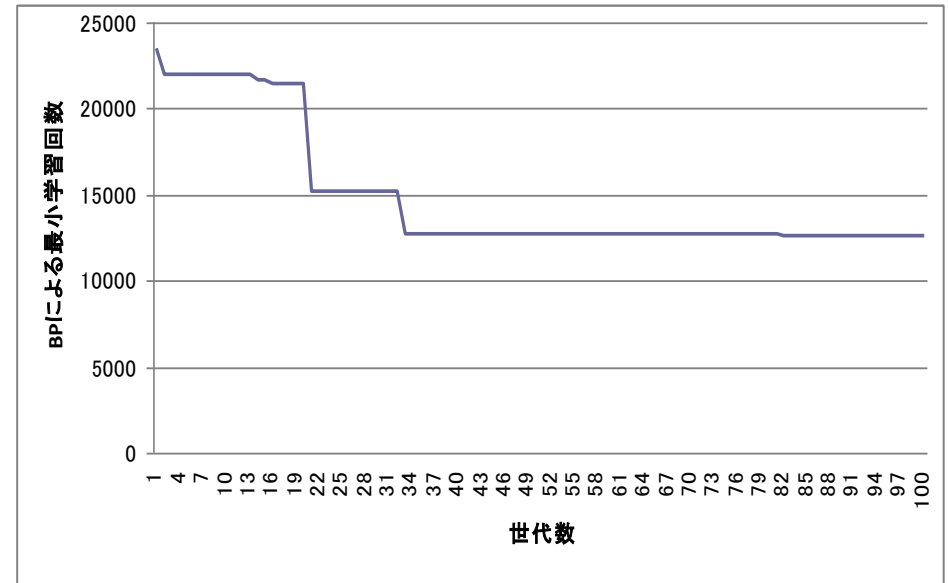


図7 実験結果
Fig.7 result of simulation

図7の実験結果は、縦軸がバックプロパゲーションによる日本語、英語の学習回数の平均値（例えば日本語が20000回で学習を終了し、英語が16000で学習を終えたとすると、 $(20000+16000) \div 2 = 18000$ となる）の最小値である。横軸はGAの世代数である。図7をみると、世代とともに学習回数の最小値が下がっているのが確認できる。このことは、BPによる個体の後天的学習とGAによる進化の相互作用により、英語と日本語の双方に共通する何らかの言語知識がエルマンネットワークの結合重みに京成されたためと見ることができる。

次に、遺伝的進化の対象となるエルマンネットワークの結合重みとして、文脈層から中間層への重みも追加することを考える。この場合、交叉として3種の結合重みの境で交叉する2点交叉を用いる。図8にその実験結果のグラフを示す。

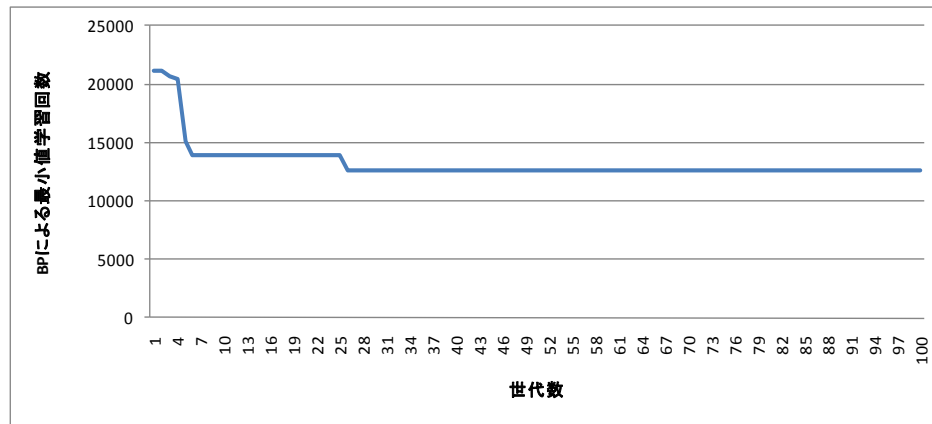


図 8 実験結果
Fig.8 result of simulation

今回も図 7 とほとんど同一の傾向が確認できる。その理由として文脈層から中間層への結合荷重に対して結合荷重に対して遺伝的アルゴリズムを適用しなくとも、隠し層 1 から中間層、そして中間層から隠し層 2 への結合荷重が進化することで、文脈層から中間層の結合荷重に対応するように変化し、学習が進むのでないかと考える。

7. 考察

シミュレーションの結果から言える事は、一部の結合荷重に対して遺伝的アルゴリズムを適用すると、日本語、英語の双方に対しての適応度の上昇が確認できたということ、つまり、日本語・英語の双方に対して、初期ランダムな結合荷重をもったエルマンネットワークよりも短時間に学習できるエルマンネットワークが遺伝的に形成されたことが確認できた。普遍文法は人間が進化の歴史の中で脳の中に獲得したものだとするならば、エルマンネットワークに遺伝的アルゴリズムを適用することにより、別々の言語に対応した二つのエルマンネットワークに対して有用な共通部分があることがわかる。このことから、普遍文法を進化的に獲得できたと考える。

二つの実験条件から行ったシミュレーションの結果をみると、初期値に関してはまったく同じではないが、実験内容はとても似ていると言える。結果を比較してみると、最小値の初期値からの減衰率や、学習回数の最小値は同じような結果である。このことから、考えられることは、文脈層から中間層への結合荷重に対して結合荷重に対し

て遺伝的アルゴリズムを適用しなくとも、隠し層 1 から中間層、そして中間層から隠し層 2 への結合荷重が進化することで、文脈層から中間層の結合荷重に対応するように変化し、学習が進むのでないかと考える。

本シミュレーションは、世代数を多くすることや、扱う言語を増やす、単語を増やすなどまだまだ多くの改良の点がある。しかし、今回のシミュレーションでエルマンネットワークに対して遺伝的アルゴリズムを適用することで、学習回数が減衰する共通遺伝子を見つけられた。これからはパラメータ調整、単語、言語の増加など改良をしていきたいと考える。

参考文献

- 1) Elman, "Distributed Representations, Simple Recurrent Networks, and Grammatical Structure, machine learning, vol.7, pp.195-225", 1991.
- 2) N. Chomsky, Lectures on Government and Binding, Dordrecht: Foris, 1981.
- 3) N. Chomsky, Knowledge and Language, New York: Praeger, 1986.
- 4) E.M.Gold, "Language identification in the limit," Information and Control, vol.10, pp.447-474, 1967.
- 5) 早瀬智英, 松田聖, "実数値 GA におけるスキーマ保存を考慮した交叉方法の提案", THE INSTITUTE OF ELECTRONICS, INFORMATION AND COMMUNICATION ENGINEERS, IEICE Technical Report AI2008-87, (2009-03).
- 6) Elman, "Finding structure in time," Cognitive Science vol.14, pp.179-211, 1990.