1A2-OS-05b-5

ディープラーニングと進化
Deep Learning and Evolution

松尾 豊[*1]
Yutaka Matsuo

[*1] 東京大学
University of Tokyo

This paper describes the relation between deep learning and evolution. Nature makes living things on earth by evolution. Some kind of insects (such as ants and bees) have eusociality: They make a society, which is realized by evolution. We human being make a society as well, which is realized by learning with our brains. In other words, evolution and learning play a similar role (although the time scale differs much) to maximize probability of ones (or kinds) to survive. Then, what is the difference between the two? This is a long-term discussion on evolutional computing. We bring a new aspect to the discussion based on recent development of deep learning.

1. はじめに

近年、ディープラーニングの技術の進展が著しい。今日のディープラーニングの研究で行われていることは、1980 年代に行われていたニューラルネットワークの研究で行われていたこととかなりの部分、重なっている。昨年 11 月に発刊された、Y. Bengio らが書いたディープラーニングの教科書[Goodfellow16]には次のように述べられている。

人工ニューラルネットワークの最初の実験が 1950 年代に行われたのに、なぜ最近になってようやく、深層学習が極めて重要な技術と認識されるようになったかは、不思議に思うかもしれない。(中略) 今日、複雑なタスクで人間の性能に到達する学習アルゴリズムは、1980 年代におもちゃの問題を解くのに苦労した学習アルゴリズムとほとんど同一である。(中略) 最も重要な新しい進歩は、今日ではアルゴリズムが成功するのに必要とするだけのリソースを、アルゴリズムに提供することができることである。

つまり、1980 年当時、できるはずだと思われていたことが、計算パワーの進展とデータの拡大により、実際にできるようになった。ディープラーニングが行っていることは、入力から出力への複雑な写像を、さまざまな prior を駆使しながら、できるだけ簡単な関数の組み合わせで、かつパラメータ数を減らすような工夫を行いながら、獲得するということである。こうしたアルゴリズムが、身体性、シンボルグラウンディング等と絡みながら、進展していくことは、[松尾 15a, 松尾 16]に述べた通りである。

翻って、人類のこれまでの歴史を考えてみると、人類は進化の過程のなかで、言葉を手に入れるだけでなく、社会を形成してきた。貨幣や宗教、法律などさまざまな社会的な仕組みを作り出し、それによって多数の人々が協力し協調する仕組みを作り出してきた[ハラリ 16]。

ところが、進化の過程そのものも、さまざまな生物における社会性を生み出している。例えば、アリやハチは真社会性をもつ生物であり、集団のなかに不妊の階級を持つ。さまざまな役割分担をして分業することで生産性を高め、種全体の維持・繁栄につとめているさまは、人間の社会とよく似たところもある。この進化で生み出された生物のなかには驚くべき社会性を発揮するものも多い。例えば、「昆虫はすごい」という本には次のような記述がある[丸山 14]。

日本の雑木林に見られるトゲアリという剌々しい大型のアリも一時的社会寄生性で、クロオオアリなどのオオアリ属のアリに寄生する。(中略) クロオオアリの巣の付近でクロオオアリの働きアリを見つけた雌アリは、その首に咬みついて動きを封じる。すると不思議な事にその働きアリはだんだんと無抵抗になる。そのように咬みついたまま、まずはクロオオアリの働きアリの匂いを自分の体に塗りつける。次に巣に侵入するが、すでにクロオオアリの匂いが体に染みついているので、攻撃を受けにくい。そしてこんどはクロオオアリの女王の首に咬みつき、長時間そのままで女王の匂いを自分の体に移す。その後、女王を殺し、自分が女王に成り代わるのである。

つまり、このトゲアリは、次々に変装をして建物に侵入する泥棒のようなことをやっている。人間の世界での泥棒は、こうした行動を、それまでの人生で学習によって得られたスキルや、言語を通して得られた知識を動員し、作戦をたてて行うわけであるが、それと同じようなことを進化は作り出したということである。

考えてみれば、生物の究極的な目的は、個体の保存ないしは種の保存であり、その目的に関して、進化も学習も同じことを異なる方法で行っていると考えることもできる。もちろん、そのタイムスケールは大きく異なり、進化は数十から数百世代をかけて徐々によくなるのに対して、学習は、個体内で数時間～数年で良くなるので、その性質は大きく違うが、人間が学習により社会性を獲得したのも泥棒を行うのも、また、アリが進化により社会性を獲得したのも泥棒を行うのも、同じような行動を獲得しているわけである。(なお、人間の学習にも当然、進化的な仕組みが相当程度使われているので、ここでの議論は単純化しすぎていることは言うまでもない。)

こうした進化と学習に対する議論は、古くからあり、例えば、ボールドウィン効果というものが議論の出発点として知られている[Depew03]。米国の心理学者ジェームズ・ボールドウィンは、1896 年に本のなかで、学習能力が種の生存能力に影響を与え、自然淘汰に効果を持つと主張した。例えば、より早く学習するものが生存上有利となり、徐々に早く学習されるようになり、あるときにそれは本能のように見える。これは、獲得形質が遺伝するというラマルク進化と似たような効果となるが、その仕組みは異なる。ボールドウィンの議論は、批判もされてきた。例えば、学習が本能になるということは必ずしも進歩とは言えない。というのも、非常に安定した環境ならば本能が役立つが、それ以外では柔

連絡先:松尾豊, 東京大学, 東京都文京区本郷 7-3-1, 03-5841-7718, matsuo@weblab.t.u-tokyo.ac.jp

軟な学習の方が優るからである。ディープラーニング研究で有名な G. Hinton もこうした議論を行っており、例えば、1987 年の論文では、学習能力を持つ個体のほうが、変化する環境における最適値を早くみつけることをシミュレーションによって示している[Hinton87]。なお、最近では、エピジェネティクスの研究が進み、獲得形質と遺伝子との関係は大変複雑であることが分かってきた。

では、こうした進化と学習の違いはどこにあるのであろうか。そして、ディープラーニングの進展は我々をどこに導こうとしているのだろうか。本稿では、この問題に対して、微分可能性、および最適化における双対問題という観点からの試論を提示する。

2. 微分可能性

進化と学習の違いはなんだろうか。

進化は、DNA によるエンコーディングを用い、それが RNA に転写され、タンパク質が合成される。DNA は、交叉や突然変異を繰り返すことにより、その環境中における「最適解」に近づく。人工知能の研究分野でも、遺伝的アルゴリズム、あるいは進化計算は、以前から主要なトピックのひとつである。

一方で、学習は、昨今のディープラーニングに代表されるように、多くのデータから最適なモデルを得る。誤差逆伝搬法が典型的だが、何らかの目的関数（尤度関数や報酬関数）を最適化するようなモデルパラメータを求めることが行われる。

この2つの本質的な違いは何だろうか。実際に、工学的な最適化のアルゴリズムとして見たときには、問題によっては、遺伝的アルゴリズムが使われる場合もあるし、ニューラルネットワークが使われる場合もある。（もちろんそれ以外の方法も多くある。）いずれも最適化問題の解法のひとつである。通常、遺伝的アルゴリズムは、メタヒューリスティックと呼ばれる組合せ最適化の解法に分類され、離散変数の最適化に用いられる。一方で、ニューラルネットワークは、他の多くの機械学習の手法と同じく、連続空間における最適化問題の解法となっている。

つまり、遺伝的アルゴリズムは、離散変数を扱うことが多く、目的関数を微分し勾配を得ることが通常は難しい。したがって、ある程度ランダムサーチに近い状態となる。（もちろん、近傍探索を行うシミュレーテッドアニーリングのような組合せ最適化の手法もある。）一方で、ニューラルネットワークは目的関数を微分し、勾配を得ることができる。すなわち**微分可能性**（differentiability）をもつ。

これは実装面の制約から理解することもできる。DNA は RNA に転写されタンパク質となるので、その上で勾配を取るというのは実質的に不可能である。生物においては通常は発達の過程を経て、さまざまな体細胞が順番に組み上がっており、DNA を変えたところで、うまくタンパク質レベルの変化に結びつかない。したがって、「作り変える」ことができず、勾配を得ることができない。一方で、学習は、進化の過程により、神経細胞という特殊な装置を見つけ出したことで、脳内の神経回路網によってソフトウェア的な手段を取り入れることに成功した。神経細胞の生成と結びつきの変化により、自由に「作り変える」ことができるようになった。したがって、勾配を取り、勾配方向に改善することができるようになった。

なお、脳の中で誤差逆伝搬が行われているのかという議論はよく行われており、これまで脳の中で誤差逆伝搬的な挙動は観測されていない。しかし、最近では、Bengio らをはじめとして、現実的な方法で脳の中でも誤差逆伝搬が実装可能なはずであると述べている[Bengio15]。

では、微分不可能なものを微分可能にすることはできないのであろうか。物理的な挙動が問題にならなければ、関数を工夫することでもちろん可能である。昨今のディープラーニングの研究では、いかに問題を微分可能にするかは大きなテーマであり、例えば、Neural Programmer という手法では、複雑な算術演算や論理推論をニューラルネットワークで表現し、end-to-end で訓練する[Neelakantan16]。深層強化学習の研究では、policy を微分可能にすることで学習を効率化する手法がさまざまに提案されている（例えば[Tamar16]）。最適化アルゴリズムのパラメータの最適化自体を、勾配をとって行おうとするアプローチもある[Andrychowicz16]。さしずめ、勾配のとれる形にさえしてしまえば、あとは深いニューラルネットワークで学習させれば何とかなるというのが昨今の風潮である。

そう考えてみると、DNA による進化というのが微分可能でないというのは少しじれったく感じる。つまり、生物の身体というハードウェアは、発達の過程を経なければならないため、エンコーディングとの関連が複雑で、勾配をとることができなかったということだろう。しかし、エンコーディングと身体の関係などの写像を数理的に記述することができ、最終的な評価関数を得ることができれば、勾配を取ることができるようになるはずである。

また、生物の身体という視点を超えて考えてみると、昨今では 3D プリンタ等の技術の進展により、ハードウェアの作り変えのコストは劇的に下がってきた。そう考えてみると、ハードウェアの適応も、勾配がとれる学習の方式に向かう（そのほうが高速に学習ができるから）ということが考えられるのかもしれない。蛇足になるが、そうした視点で見ると、2016 年に公開された映画「シンゴジラ」は、ハードウェアの作り変えを個体内で（おそらく学習により）行っており、実は生物の機能としてより進んだものであり、妥当な描写かもしれないと感じた。

3. 生存確率の最適化と双対性

前節では、微分可能性が重要であるということを議論した。本節では、目的関数について議論する。

我々が生きている世界における生物の目的は、自己の保存、あるいは種の保存であろう。すなわち、同一性の保存といえるだろう。なお、同一性とは何かということ自体、認知現象と関わるため、さまざまな問題を孕んでいることは[松尾 15b]に述べた。しかし、同一性の定義がどうあれ、世界には、自己保存するように行動したものが、結局は自己保存するという単純な法則しかないようにも思える。

すなわち、生物は、**生存確率の最大化**をしていると見ることができる。このとき、単純化したモデルを考えると、環境中からなんらかの情報 $o \in O$ を観測し、それを処理し、行動 $a \in A$ を行うことで、生存確率を最大化する。当然、情報 o の取得にはコストがかかるし、行動 a の選択にはコストがかかる。生物として、取得可能な o の集合 $O' \subset O$ を用意し、また、選択可能な行動の集合 $A' \subset A$ を用意していると考えよう。生物でいうと、O'や A'を選ぶ行為は、センサやアクチュエータ、すなわち身体の形状を選ぶことに相当し、主に進化で行われてきた。そして、o から a につなぐ処理は、何らかの情報処理によって行われる。この部分は、原始的な生物ではほとんどがハードウェアによって、哺乳類等では、かなりの部分が（進化的な仕組みで誘発された）学習によって行われている。この処理は、重みの最適化だけでなく、表現の獲得も含み、ディープラーニング、深層強化学習、さらには、その上にのるべき確率モデル、言語モデル、コミュニケーションのモデル等になる。

一方で、通常、最適化問題には**双対問題**が存在する。では、生存確率の最大化を主問題としたときに、双対問題にあたるものは何であろうか。双対問題にもさまざまなタイプがあるが、ラグランジュの双対問題や Wolfe の双対問題などでは、どの制約が

アクティブになっているか自体を変数にとる。主問題に照らして言うと、どの情報 $O' \subset O$ を観測するか、そして、どのアクション $A' \subset A$ を用いるかが変数である。つまりは、センサやアクチュエータの設計である。そして、o から a への写像において、どの制約がアクティブになっているか、すなわちどういった**表現**が使われているかが変数になる。(詳しくは付録を参照のこと。)

そして、最小化すべき関数は、こうしたセンサやアクチュエータの選択、表現の獲得を無視した場合の生存確率から、制約の違反度をペナルティとして足し合わせたものになる。つまり言い方を変えると、生存確率の最適化に必要な情報のロスを最小化しているとも言える。

こう考えると、次のように言える。

1. 主問題で対象としているのは、生存確率の最大化である。主変数は、センサの情報やアクチュエータの情報、どのようにセンサ情報をアクチュエータの情報に変換するか(例えば、ニューラルネットワークの重み係数)などである。
2. 双対問題で対象としているのは、生存確率を高める上で必要な情報のロスの最小化である。双対変数は、どのセンサを使うか、どのアクチュエータを使うか、そして、どの表現を使うかなどになる。

つまり、大雑把に言うと、主問題では生存確率の最大化をするために、観測した情報からどのように行動を学習するかを扱っている。双対問題では必要な情報のロスを最小化するために、そもそもどのような情報を観測し、どのような表現を用いるのかを扱っている。**物理世界における最適化が主問題**であり、**情報世界における最適化が双対問題**ということもできる。

ディープラーニングで行っている学習は、主変数である重みやバイアスの学習と見ると、主問題を解いていることになるが、そうして得られた表現を獲得していると考えると、主問題ではなく双対問題を扱っていることになる。これが、従来の機械学習とディープラーニングの本質的な違いが直感的に理解されにくい理由であるかもしれない。そして、表現の問題は、センサの選択やアクチュエータの選択、環境とのインタラクションという問題も広く含む[松尾 15b]。

そういった意味では、これまでの地球上の生物の歴史的な進展は、あるときには主問題から、あるときには双対問題からアプローチしてきたと言えるのかもしれない。特に、双対問題からのアプローチ、特に表現の選択に関する部分が長らく難しかったのは、情報の保持が難しかったということだろう。人類はこれを言語や社会制度によって解決した。

言語による参照能力は、非常に大きな変革をもたらした。任意のものを言語により参照できる機能は、無意識のさまざまなプロセスと独立に動くことができる。この言語によるプロセスで最も強力なものが、情報を蓄積しておく能力であろう。内的な言語により神経回路の情報を抽象化し、情報量を減らすことで、それを長期に蓄積しておくことができるようになった。(なお、意識と言語を関連させる議論は多いが[甘利 16]、言語による自分自身への参照が意識の感覚を産むのかもしれない。)

また、それを外的な言語と結びつけることで、他者とのコミュニケーションに用いることができるようになり、複数のエージェントをまたがる情報交換を可能にした。また貨幣は、文字通り「報酬」として使われ、強化学習における報酬と同様の機能を果たすことによって、人間の集団全体において環境に適応するためのサブゴール(あるいは微分可能性)を細かく与えることとなった。さらに法律は、個体の動きを制限し、集団が目的となる行動に近づくことを可能にした。経済学者の岩井克人氏は、言語、貨幣、法律を3つの重要な発明に挙げている[岩井 06]。そして、近年では、インターネットの出現、検索エンジンの出現により、情報の取得の可能性は大きく向上した。つまり、言語や貨幣、インターネットによって、双対問題の解き方が、個を越えて集団レベルに変化したということであり、それにより、双対問題からのアプローチがますます優勢になってきたであろう。

こうして考えると、生物の生存という最適化問題に対して、従来は物理世界を対象とする主問題からのアプローチが主であった。しかし、人類は、学習の仕組み、言語や社会などの仕組みを作り出すことで、微分可能性を上手に活用しながら、情報世界をメインとする双対問題からのアプローチにスイッチし、そして結果的にこの最適化問題の最適値を大きく更新してしまったということができるのかもしれない。

4. 微分可能化する世界

では、今後はどのような変化が起こるのだろうか。先に述べたように、生物の生存確率の最適化において、DNA という仕組みを用いることによるハードウェアの改善の難しさに起因して、勾配を取るのが難しいという問題があった。ところが、それが簡単に作り変えられるようになる世界では、双対問題における特にセンサやアクチュエータを含めた最適化の能力が格段に上がるはずである。

シンギュラリティ論にうかつに立ち入るのはあまり適切でないと考えているが、ここでは、そうした先のことまで想像してみよう。人間が作った人工知能が全人類の知能を超えるとか、あるいは人工知能が人工知能を生み出すような、シンギュラリティで描かれている世界というのは、それほど本質的な面を捉えているとは思わない。一方で、今後、数十年〜数百年のうちに、例えば遺伝子工学の進展により、人間は死ななくなるのかもしれないが、結局は、こうした最適化問題の解き方がより効率的になると考えるほうが自然なのではないだろうか。つまり、何らかの「自己同一性」を保存する確率の最大化を行い、また、同時にそのために必要な情報量の最大化を行うわけである。そうした先には、ほぼ確実に自己同一性が失われず、重要な情報の欠落がほとんどない状態に近づくことになる。

そうして考えると、私は、「**世界は微分可能化している**」と強く感じる。微分可能にしたほうが明らかに改善の速度が速い。ディープラーニングにおける研究は、いかに微分可能な関数を定義し、データを用いて訓練するかを競っている。一方で、これまで、社会制度や組織、社会慣習などは微分可能ではなかった。しかし、昨今では、例えば、企業活動をデータ化し、ビッグデータの分析により改善しようという試みが多く行われる。通常は何らかの KPI を定めて、これを改善するような施策をとる。これは、与えられた KPI に対しての勾配を求めていることに他ならない。そうすると、従来は経営者の勘と経験に頼っていた部分がより効率化されるわけである。例えば、ウェブサイトの最適化や設計をどこまで自動で行えるかという研究[Iitsuka15]も行われているが、世界は確実に微分可能な方向に向かっているのではないだろうか。そして、我々の社会で作り出す製品、サービス、そして社会システムまでもが、どこまで微分可能になるのであろうか。

なお、念のため付記しておくと、こうした世界観はディストピアではない。何も人間という存在がいなくなる/いらなくなるのではなく、人間が言語や貨幣を発明することが、期せずして双対問題からの解法を増強したように、結果的に、人間と人工知能を含む社会全体が、こうした極限状態に近づくということである。現代社会においても、人々は夜眠り、朝日や夕日を美しいと思う。過去から蓄積された本能、感情の上に、我々の社会が構築されており、それは人間あるいは生物の「生きたいと願う本能」に大きくドライブされている。そうした生物としての本能・感情がドライバとなり、主問題・双対問題の解法がより進展するわけで

ある。また、ここで議論しているような世界観は、かなり長い時間スケールでの変化を敷衍したものであり、そのリアリティに関しては、巷の「シンギュラリティ論」と同程度に空想上のものであると断っておきたい。

5. おわりに

本稿では、ディープラーニングの進展からはじまって、それが進化とどのような関係にあるか、双対性の観点から議論した。進化と学習は、古くからある難しいテーマであるが、ディープラーニングと関連させて議論した。さらに、微分可能性の導入による進展の加速、その先にある世界観についても提示した。

ここで述べたことは、かなり空想的なものであるが、多少なりともこの世界についての本質を捉えているのではないかと期待する。そして、いつも述べていることであるが、よりリアリティをもった話として、日本がここでいう双対問題における「ハードウェアの微分可能性」という産業的なチャンスをものにして欲しいこと、そして、人間の（知能を除いた）人間性に関する本質的な考察が、人文社会学の学問体系を再構築する形で生まれてきて欲しいこと、こうした研究がより良い人類社会の実現につながって欲しい[倫理 17]ことを述べて、結言としたい。

参考文献

[Andrychowicz16] M. Andrychowicz et. al, Learning to learn by gradient descent by gradient descent, Proc. NIPS2016, 2016

[甘利 16] 甘利俊一: 脳・心・人工知能　数理で脳を解き明かす, 講談社, 2016

[Bengio et al. 15] Y. Bengio, D. Lee, J. Bornschein, T. Mesnard, and Z. Lin, Towards Biologically Plausible Deep Learning, arXiv:1502.04156, 2015

[Depew03] D. Depew, Evolution and Learning: The Baldwin Effect Reconsidered, MIT Press, 2003

[Goodfellow16] I. Goodfellow, Y. Bengio, and A. Courville, Deep Learning, MIT press, 2016

[ハラリ 16] ユヴァル・ノア・ハラリ (著), 柴田裕之 (翻訳): サピエンス全史 文明の構造と人類の幸福, 河出書房新社, 2016

[Hinton87] G. Hinton and S. Nolan, How Learning can Guide Evolution, Complex Systems, No.1, pp.495-502, 1987

[Iitsuka15] S. Iitsuka and Y. Matsuo, Website Optimization Problem and its Solutions. Proc. KDD 2015, 2015

[岩井 06] 岩井克人, 三浦雅士:資本主義から市民主義へ, 新書館, 2006

[カーツワイル 07] レイ・カーツワイル:ポスト・ヒューマン誕生 コンピュータが人類の知性を超えるとき, NHK 出版, 2007

[丸山 14] 丸山 宗利:昆虫はすごい, 光文社, 2014

[松尾 15a] 松尾 豊:人工知能は人間を超えるか ディープラーニングの先にあるもの, KADOKAWA/中経出版, 2015

[松尾 15b] 松尾 豊:Deep Learning と人工知能の発展, 人工知能学会全国大会, 2C3-OS-06b-5, 2015

[松尾 16] 松尾 豊:Deep Learning から身体性、シンボルグラウンディングへ, 人工知能学会全国大会,1A5-OS-27c-2, 2016

[Neelakantan16] A. Neelakantan, Quoc V. Le, and Ilya Sutskever, Neural Programmer: Inducing Latent Programs with Gradient Descent, Proc. ICLR 2016, 2016

[Tamar16] A. Tamar, S. Levine, P. Abbeel, Y. Wu, and G. Thomas, Value Iteration Networks, Proc. NIPS2016, 2016

[倫理 17] 人工知能学会 倫理委員会:倫理指針, http://ai-elsi.org/, 2017

付録

本文で議論している主問題は、例えば次のような形を取る。

$$\text{maximize } E[P_e(o, a)]$$

$$\text{subject to } a = f(o),\ a \in A',\ A' \subset A_e,\ o \in O',\ O' \subset O_e$$

ただし、P_eは、環境 e において、ある主体が、観測 o が得られたときに行動 a をとるという場合の生存確率であり、その期待値を最大化している[1]。A_e は、環境 e において取り得る行動の集合、O_e は観測し得る情報の集合であり、f は o を a に写像する関数である。観測 o や行動 a は次のように定義されるとする。

$$0 \leqq o_1 \leqq o'_1,\ o'_1 \geqq 0$$
$$0 \leqq o_2 \leqq o'_2,\ o'_2 \geqq 0$$
$$\ldots$$
$$0 \leqq a_1 \leqq a'_1,\ a'_1 \geqq 0$$
$$0 \leqq a_2 \leqq a'_2,\ a'_2 \geqq 0$$
$$\ldots$$

つまり、o'_i がゼロであれば、o_i もゼロとなる。つまりそのセンサは使われない。o から a に写像する関数 f はさまざまなものがあるが、例えば、o に対して何らかの計算をし、階層的な n 層から成る中間表現 h を作るとすると、次のようになる。

$$h_1 = g_1(o)$$
$$h_2 = g_2(h_1)$$
$$\ldots$$
$$a = g_n(h_{n-1})$$

このような h は無限にあり得るので、どのような h を使うかは、表現の使用可能性を表す変数 h'によって制限されているとする。

$$-h'_1 \leqq h_1 \leqq h'_1,\ h'_1 \geqq 0$$
$$-h'_2 \leqq h_1 \leqq h'_2,\ h'_2 \geqq 0$$
$$\ldots$$

主問題に使われる主変数としては、$a,\ a',\ o,\ o',\ h,\ h'$, あと、学習時に学習されるパラメータである例えば重み w やバイアス b などとなる。より複雑な処理になると、他にも多数のパラメータが加わることになる。主変数をまとめて x と書く。

一方で、双対問題は、以下になる。

$$\text{minimize}_{\lambda}\ \sup_{\lambda, x}\ E[P_e(o, a)] - \Sigma\lambda_{o'} o' - \Sigma\lambda_{a'} a' - \Sigma\lambda_{h'} h' - D$$

λ は $o',\ a',\ h'$の非負制約に対応する非負の双対変数、D は、それ以外の制約に対する項をまとめたものである。この関数は、センサやアクチュエータの選択、表現の選択を無視した場合の期待生存確率から、制約の違反度をペナルティとして足し合わせたものになる。

主変数におけるセンサやアクチュエータの制約（例えば $o'_1 \geqq 0$）がアクティブになったとき（すなわち $o'_1 = 0$）のときに、この制約に対する双対変数 λ_{o1}は正の値を取る。つまり、使われていないセンサ・アクチュエータに対して、対応する双対変数は正の値を取る。したがって、双対変数は無数にあることになり、双対問題を解くほうが、主問題を解くより通常は難しいと思われる。使われるべきセンサ・アクチュエータが使われていないときには、双対変数は正の値を取り、双対問題のラグランジュ項も正の値を取り、主問題と双対問題の解は一致しない。

主問題と双対問題の解は、最適解においては一致する。このとき、主体の生存確率は最大化され、必要な情報のロスはなくなっているという状態に相当する。

[1] 通常の強化学習では、時系列の o を加工した状態 s を取るが、ここでは単純化している。ここでの定式化は非常に単純化したものであり、状態だけでなく、センサやアクションのコスト、そもそもセンサやアクションの定義域が定義できるのか、インタラクションによる環境の改変など、重要な多くの要素を捨象している。