

遺伝的プログラミングによる物理法則の発見

Discovery of the law of physics based on Genetic Programming

松原 恒介 (知的システムデザイン研究室)

Kousuke MATSUBARA (Intelligent Systems Design Laboratory)

Abstract Genetic Programming is attracting the attention as one of data mining methods. GP is a method that extends gene expression to tree to handle sign and can evolve programs. But the study of discovering physical laws from immense observation data using GP doesn't do. So in this study, I experiment of discovering physical laws using GP from observation data that contains immense and much noise. Here, I adopt "equal energy laws" as one of physical laws. First, from equal energy law I make artificial data that doesn't contain noise, and experiment. Next I actually experiment device, and experiment to measure observation data. Finally, I compare and examine each experiment result.

1 はじめに

データマイニングあるいはデータベースからの知識発見 (Knowledge Discovery in Database) は、データベースに蓄積されたノイズを含む膨大な生データから、高いレベルで記述された価値ある情報を発見する研究であり、活発に研究が行われている [1][2]。このようなノイズを含む生データから法則を発見する分野の一つとして、物理法則があげられる。未知の体系の実験データから系の特性を表現する普遍的な関係式を求めることは科学的法則発見と呼ばれている。しかし、計算機による法則を自動発見する試みの多くは単なる実験式・経験式の候補を多数導出するレベルに留まり、複雑な対象について普遍性を有する本来の意味での法則を発見できる実用的レベルには程遠い [3]。ここで中野らは、ニューラルネットを用いたアプローチによって膨大な観測データから物理法則を発見している [4]。

一方、データマイニングの手法の一つとして遺伝的プログラミング (Genetic Programming, 以下 GP) が注目を集めている [5][6][7]。GP は生物の進化を模倣した遺伝的アルゴリズム (Genetic Algorithms, 以下 GA) の考え方に基いて、記号的な処理ができるように解の表現を木構造に拡張してプログラムなどの進化を可能にした手法である [8]。しかし、GP によって膨大な観測データから物理法則を発見する研究はされていない。そこで本研究では、データマイニングによって膨大かつ多数ノイズを含む観測データから、GP によって物理法則を発見する実験を行う。ここでは物理法則の一つとして「力のつり合いの法則」を取り上げる。まず、力のつり合いの法則からノイズを含まない人工データを作成し、実験を行う。次に実験装置を作成し、観測データを測定することで実験を行い、それぞれの実験結果を比較検討

する。

2 遺伝的プログラミング

GP は GA を拡張させたもので、John Koza によって提唱された手法である [8]。この GA と GP にはいくつか類似点があげられる。両手法は集団における個体として表現された多くの解の集合によって構成されている。これらが世代と呼ばれるサイクルによって、与えられた個体を遺伝的オペレータと呼ばれる、交叉、再生産、突然変異を行うことで新しい個体を生成する。また、これらの遺伝的オペレータは適合度と呼ばれる評価値によって確率的に実行される。この適合度は評価関数と呼ばれる式を生成する。また、この適合度は評価関数と呼ばれる式を生成する。また、この適合度は評価関数と呼ばれる式を生成する。

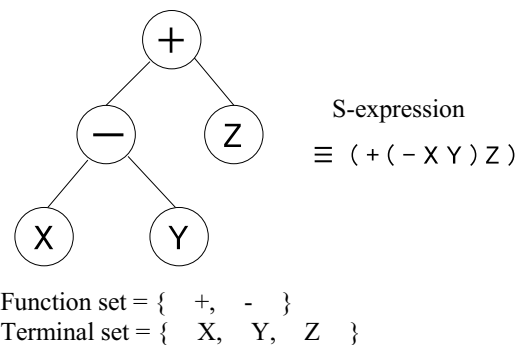


図 1: 木構造

また、GA と GP の相違点として GP では個体のもつデータ構造が木構造であることがあげられる。図 1 は木

構造の一例で $(X-Y)+Z$ を表現している．ここで子を持つことのできる記号 $+$, $-$ といったノードを非終端記号という． $+$, $-$ の両ノードとも 2 つの子を持つため 2 つの引数を持つという．また記号 X , Y , Z といった子 (引数) をもたないノードを終端記号といい、木構造の末端に配置される．図 1 に示す表現は LISP で用いられる S 式と呼ばれ、木構造と同一視される．GP ではこのような S 式の木構造によって求める関数を表現する．

3 使用するつり合いのモデル

ここで実際に訓練データ、テストデータを与えるためのモデルについて説明する．図 2 は 3 力のつり合いを表したもので、おもりとして P_1, P_2, P_3 を使用し、これらは糸により中央の点 $Q(Q_x, Q_y)$ でつながっている．また、それぞれのおもりは台の端でつり下げられ、それぞれの端部分を $P_1(X_1, Y_1), P_2(X_2, Y_2), P_3(X_3, Y_3)$ とする．また台の大きさは縦 \times 横が 10×10 の正方形であり、座標は台の左下が $(0, 0)$ 、右上が $(10, 10)$ とする．このつり合い状態でのおもりの重さ、それぞれの座標を学習データとして与えることでつり合いの関係式を求める．このように 3 つのおもりがつり合っている状態で、 P_1, P_2, P_3 から中点 Q にかかる力をそれぞれ F_1, F_2, F_3 としたとき、次のような関係があげられる．

$$\sum_{i=1}^3 F_i = 0 \quad (1)$$

この式が力のつり合いの法則である．この式を、おもりの重さおよび座標によって変形すると次の式になる．

$$\frac{3}{P_1(O, Y_1)} \quad (2)$$

こ 作成
し、

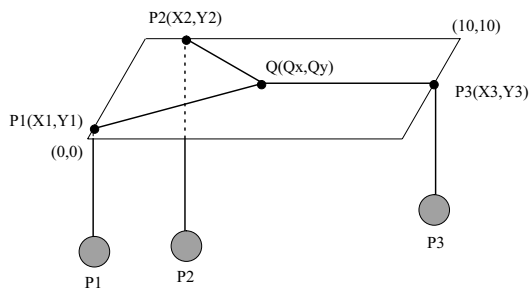


図 2: つり合いのモデル

4 人工データおよび GP パラメータ

本実験では、表 1, 2, 3 の GP を実行することで最も誤差の小さな学習データを調査する．ここでは、 $X_1, X_2, X_3, Y_1, Y_2, Y_3$ が固定されたモデルを 5 変数モデル、可変のモデルを 11 変数モデルと呼ぶ．

表 1: 3V_5

	3V_5_Qx	3V_5_Qy
入力データ	P1,P2,P3	P1,P2,P3
出力データ	Qx	Qy
モデル	5 変数	5 変数

表 2: 3V_11

	3V_11_Qx	3V_11_Qy
入力データ	P1,P2,P3	P1,P2,P3
出力データ	Qx	Qy
モデル	11 変数	11 変数

表 3: 10V_11

入力データ	P1, P2, X1, X2, X3 Y1, Y2, Y3, Qx, Qy
出力データ	P3
モデル	11 変数

以上のような関係をもつ訓練データおよびテストデータを作成し、3V_5, 3V_11, 10V_11 における GP の評価実験を行った．また、GP におけるそれぞれの非終端記号は、数学的初等関数である ($+$, $-$, $*$, $/$, \sin , \cos , \exp , \log) の 8 種類を用いた．選択方法としては、tournament 選択を用いた．適合度計算は、予測値 (dv) と理想値 (v) の差である $dv-v$ を用いた．

5 実験結果

ここでは GP によって求められた予測値をつり合いの法則に当てはめ、力のつり合いの誤差を求める．ここで使用される GP のパラメータを表 4 に示す．また交叉率は 0.9 から 0.5, 突然変異率は 0.0 から 0.4, 再生産率は 0.1 とした．

このパラメータによって 3V_5_Qx, 3V_5_Qy, 3V_11_Qx, 3V_11_Qy, 10V_11 を lil-gp[9] によって実行した．それぞれの実験でのつり合いの誤差で最良の結果を表 5 に示す．

実験結果から最も誤差の少なかったのは 3V_5 である．これは 10V_11 では終端記号が多いため誤差が大きくな

表 4: GP のパラメータ

データ数 (訓練 , テスト)	(100 , 100)
世代数	100 世代
個体数	10000
最良個体の選択方法	テストデータ
適合度関数	dv-v
非終端記号	8 種類

表 5: 誤差の比較

	訓練データ	テストデータ	パラメータ
3V_5_Qx	9.6637	11.4933	0.8,0.1,0.1
3V_5_Qy	15.8428	16.7150	0.5,0.1,0.4
3V_11_Qx	145.9261	158.5731	0.7,0.1,0.2
3V_11_Qy	377.8859	324.3468	0.8,0.1,0.1
10V_11	43.5133	44.6085	0.6,0.1,0.3

り, 3V_11 では関数を求めるために必要な変数が足りなかったため誤差が大きくなったと考えられる。ここで最良の結果での予測値と実際の値を比較するグラフを図 3 で示す。グラフの X 軸における 1 から 30 は訓練データによる比較, 31 から 60 はテストデータによる比較である。グラフより訓練データによる結果ではほぼ近い関数を求めることができ, テストデータによる結果でもほぼ近い値を求めることができた。よって, 力のつり合いを求める GP の入力データは P1, P2, P3, 出力データは, Qx あるいは Qy をもつ 5 変数モデルの関数とする。3V_5 の入出力データによって実験装置を作成し, ノイズの含まれる観測データによって GP の実験を行う。

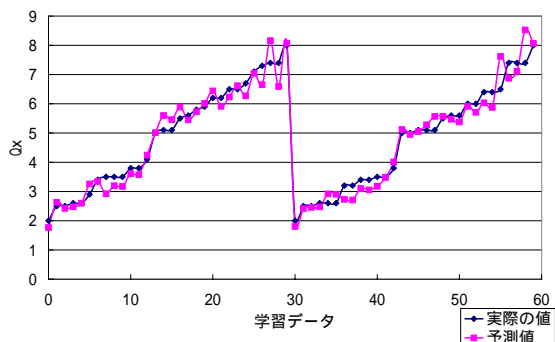


図 3: 予測値と実際の値の比較

6 結論

本研究では, GP によって物理法則の一つである力のつり合いの法則を導出できるかの検証を行った。この初段階として, 人工データを作成し, GP に適用した。結果として, 5 変数のモデルで 3 つのおもりの重さから中点を求める関数を求めることができた。また, この関数は誤差の小さな高品質な解であった。今後として, この 3 力のつり合いの実験装置を作成する。この装置からデータを観測し, GP に適用する。この実験より, 摩擦などノイズを含む環境条件の下での GP の精度について調査する。

参考文献

- [1] 河野浩之:データベースからの知識発見の現状と動向, 人工知能学会誌, Vol.12, No.4, pp.497-pp.504, 1997
- [2] 山本英子, 梅村恭司訳:データマイニング, 共立出版, 1998
- [3] C. Schaffer: Bivariate scientific function finding in a sampled, real-data testbed, Machine Learning, 12(1/2/3), pp.167-pp.183, 1993
- [4] K. Saito and R. Nakano: Law discovery using neural networks, Proc. of 15th International Joint Conference on Artificial Intelligence (IJCAI '97), pp.1078-1083, 1997
- [5] 田淵真, 田浦俊春:遺伝的学習機構と人との対話型知識獲得手法, 人工知能学会誌, Vol.11, No.4, pp.600-pp.607, 1996
- [6] Peter J. Angeline and Kenneth E. Kinneer, Jr.: Advances in Genetic Programming Volume 2, Massachusetts Institute of Technology, 1996
- [7] M.L. Raymer, W.F. Punchi, E.D. Goodman, L.A. Kuhn: Genetic Programming proceeding of the First Annual Conference, MIT Press, 1996
- [8] 伊庭斉志:遺伝的プログラミング, 東京電機大学出版局, 1996
- [9] isl.msu.edu/GA/software/lil-gp/index.html