人間のヒューマノイド操縦上達過程における 感覚運動制御の変化に基づく評価器自動生成法

Automatic Generation of Evaluation Function from Human's Improvement Process
of Intuitive Operation of Humanoid Robot based on time series of Sensor-based Motion Control

○鈴木駿介(農工大) 湯本裕矢(農工大) 正 水内郁夫(農工大)

Shunsuke Suzuki, TUAT, ssuzuki@mizuuchi.lab.tuat.ac.jp

Yuya Yumoto, TUAT

Ikuo Mizuuchi, TUAT

Various complex motions such as skillful dynamic motions have been achieved by various researchers. For generating such motions, machine learning has been utilized. However, it is difficult to design an appropriate evaluation function for complex motions such as smashing roof tiles and turning kick. We focus on human's ability of improving various kinds of sensor-based motions. It seems that a human can generate evaluation functions for complicated sensory-motor coordination during the gradual improvement process. This paper proposes a method of emulating human's evaluation functions based on the improvement process of operating a humanoid by a human. The human will be good at operating the humanoid based on displayed sensory information of the humanoid such as ZMP sensors and IMU. We tried a recurrent neural network for imitating human's subjective evaluations of each motion. This paper indicates that we could automatically generate evaluations. In addition, we compared the evaluation function with other evaluation function of machine learning and proved the validity of the evaluation function during the improvement process of operating a humanoid by a human.

Key Words : Humanoid, Evaluation Function, Recurrent Neural Network

1. 序論 -人間の上達過程を基にした評価関数自動生成-

ロボットはこれまで走行や跳躍など様々な運動を実現している．中でも実現する方法がわからない動作の生成や，個々の動作を作る手間を省くための手段の一つとして機械学習という方法がある^[1]．機械学習の利点は，適切な評価関数を定めることで実現方法のわからない運動にも適用が可能であることである．しかし，それぞれの運動の評価関数を定めるには経験や勘に頼って試行錯誤する必要がある，個々の動きにしか対応出来ない等の課題があった．それぞれの運動に関して，評価関数を自動で生成出来れば，人間が評価関数を逐一試行錯誤して作る手間を省くことが期待できる．

そこで本研究では，様々な運動を実現可能な人間の上達過程に注目した．人間の上達過程において対象の運動に関して，何に注目し，何を修正したかを解析することが出来れば対象の運動の評価関数の生成の仕方が判り，更には評価関数の自動生成が可能になるのではないかと考えた．本論文では，人間の試行錯誤実験による上達過程を基に評価関数の自動生成を行い，運動の評価を行った．また，人間の上達において生成した評価関数と機械学習で使用する評価関数との比較を行い，自動生成した評価関数の妥当性について検討した．

2. 人間の上達過程における評価関数生成

人間は，実現の仕方が判らなかった運動を，何度も試行錯誤を重ねて上達している．ここで，人間が上達しているうちに無意識に評価関数を脳内で徐々に生成していると仮

定する．このとき，どのような評価関数が具体的に判らなくても人間は実現の仕方の判らなかった運動を実現出来ている．本研究では，図1のようにヒューマノイドの操縦を用いて動作を上達し，その過程からヒューマノイドの動作の評価関数の自動生成を行い，人間が評価関数を作る手間を省けるようにする．

3. 人間の試行錯誤実験の環境構築

3.1 ヒューマノイド操縦システムの構築

ヒューマノイドは二腕二脚，多自由度(本研究で使用するヒューマノイドは25自由度)であるので，人間が操縦するにはジョイスティック等の少数自由度インターフェイスでは難しい．また，操縦する人間の上達過程をヒューマノイドに活用する場合，人間の動きを出来るだけ直観的に伝達できることが望ましい．本研究では Microsoft 社製の

Fig.1 System of improvement process of operating a humanoid robot

Fig.2 System of observing the state of a humanoid

Kinect for Xbox360 を用いて人間の動作をモーションキャプチャして、その情報から逆運動学を解いてヒューマノイドの操縦を行う。

3.2 ヒューマノイド状態観測システムの構築

試行錯誤実験において操縦者は、ヒューマノイドを視覚で観察せずにヒューマノイド状態観測システムのみを用いて観測し、実験を行う。これは、操縦者へ提示する情報を限定することで、人間の上達にとってどのような情報が有用であるかを判断しやすくするためである。

ヒューマノイド状態観測システムには図2のようにディスプレイにヒューマノイドの足先位置および重心位置、Zero Moment Point(以下 ZMP)^[2]、ヒューマノイドの簡易モデルを表示する。

4. 動力学シミュレータを用いたヒューマノイド操縦実験

4.1 実験方法

本研究では、客観的に良いかどうかを判断しづらい運動の評価を行う。客観的に評価しづらい運動の例として、瓦割り運動と回し蹴り運動の2種類の運動に着目した。瓦割り運動では、単に腕を振って瓦を割るだけでなく美しいフォームを目指して運動を行う。美しいかどうかは主観的に判断し、評価を行う。このように、運動の美しさのような定量的に数値化しづらい対象を評価し、数値化出来るようにすることを提案する。回し蹴りでは足先の撃力や速度、全身の運動フォームの美しさに注目し、運動を行う。それぞれの運動の試行実験を1試行につき40秒間行う。

4.2 実験結果

図3は、それぞれ瓦割り運動と回し蹴り運動の評価結果である。瓦割り運動実験は200回、回し蹴り運動実験は60回行った。図4に瓦割り運動の199回目の動作の様子を、図5に回し蹴り運動の55回目の様子を示した。それぞれの運動の実験結果は、主観的に最もよかったものを1.0として、1回の試行毎に評価値をそれぞれ比較して0.0~1.0の範囲で0.1刻みで評価をつけたものである。図3を見る

Fig.3 Result of evaluation output

Fig.4 The experiment of 199th smashing roof tiles

Fig.5 The experiment of 55th turning kick

と後半の試行に高評価が多い傾向にあり、操縦者がそれぞれの運動を上達出来ていると言える。

5. 試行錯誤実験データによる評価器生成と運動の評価

人間の上達過程におけるヒューマノイドから得られるZMP等の数値の変化から、対象の運動に対しどの要素が重要で、どの要素があまり重要でないかを定めるのは困難である。そこで、実験した運動に対し、ヒューマノイドから得られるZMP等のベクトル量を入力するとその運動の評価を出力するような評価器の生成手法を提案する。このような評価器を生成することが出来れば、運動を評価することが出来たことになる。更には、この評価器生成によって、美しさなどの定量的に評価できない対象を定量的に評価出来ることが考えられる。ベクトルデータは時系列になっており、時系列ベクトルを扱う方法の一つとしてリカレントニューラルネットワーク(以下RNN)がある^{[3][4]}。図6の上半分のようにRNNを用いて評価関数を生成し、下半分のように生成した評価器を用いて運動の評価を行う。

5.1 使用するネットワーク構造の決定

本研究では、RNNの一つであるBackpropagation Through Time法(以下BPTT法)^[5]を用いることにした。BPTT法は時刻 t_0 から t までのネットワークの入出力計算を行い、そのときの入出力と重み係数を記録し、その後、 t から t_0 までの逆時間で誤差の計算、および重み係数を更新する。また、今回使用する層は図6にもあるように3層とする。このとき、入力層のユニット数はベクトル

Fig.6 Construction of evaluation function generator by using RNN

時系列の次元数とフィードバックによる次元数の和であり、出力ユニットは試行者のスキルの評価値となるため、次元数は1となる。

入力層の数は31次元の時系列データ及びフィードバックによる1次元の和(計32)とする。入力には、現在関節角度や重心加速度、全身姿勢角度、足裏感圧センサ値等を用いる。このとき、入力は150フレームの時系列データとする。このフレームは瓦割りでは腕を振り下ろし終えたとき、回し蹴りでは蹴り終えたとき、または転倒してヒューマノイドが床に着いたときを終フレームとして150フレーム分を扱うことにする。フレームレートは20~40msecである。

しかし、中間層のユニット数は以上のような条件では決定することはできない。ユニット数を客観的に決める方法の一つとして、Rissanenが提唱した最小記述長 (Minimum Description Length 基準, MDL) ^[6] を用いて定めた。

$$MDL = \frac{1}{2} N k \log(\sigma^2) + \frac{M}{2} \log(N) \quad (1)$$

このとき、 N は学習データ数、 k は出力層のユニット数、 σ^2 は出力誤差の分散値、 M は時系列のパラメータ数である。これにより数値を最小にするユニットがもっとも適切とされる。また、それぞれの時刻に対し、教師との許容誤差がある一定値に達するまで、RNNをループさせる。許容誤差範囲内に収束しない場合に備え、ループ数に限界値を設けた。

5.2 RNNによるテスト評価

人間が操縦したときのデータに関して、教師をうまく再現できた組み合わせを基にRNNで複数のデータの評価テストを行い、教師ではないデータも評価できるかどうかを検証する。教師データが2~4種類の場合で評価テストを

Fig.7 Result of RNN test for two supervisor data

Fig.8 Result of RNN test for three supervisor data

Fig.9 Result of RNN test for four supervisor data

行った。このとき、教師が5種類以上では、操縦データの教師を再現することが出来なかったため、5種類以上では検証しないことにする。教師データはそれぞれ評価の異なる似ていないデータである。教師が2種類、3種類、4種類の結果を図7、図8、図9に示す。グラフの横軸・縦軸はそれぞれ人間による評価値とRNNによる評価値である。凡例に示してある数値は教師データの人間による評価値である。また、黒線は人間の主観的な評価値とRNNによる評価値が等しくなる場合の理想的な線であり、それぞれのテスト評価が黒線に近いほど人間の主観的な評価をRNNは再現することができたことになる。

それぞれの図を比較しても判るように、教師データが増えるにつれてRNNによる評価値が黒線に近づき、正確さが増えていることが判る。このことから、RNNによる評価器は、似ていない教師データの数が多くほど、人間の主観的な評価を再現出来る傾向にあることが判った。

Fig.10 Smashing roof tiles motion acquired with evaluation function for hand velocity maximization

Fig.11 Evaluation result of smashing roof tiles motion that is generated by motion learning with evaluation function for hand velocity maximization

6. 異なる運動生成法による運動評価と人間の主観的評価の比較

生成した評価器を用いて、本研究と異なる運動生成手法で得られた運動を評価したとき、どのように評価されるか検証した。今回は湯本らの研究によって生成した瓦割り運動の学習後の時系列データを評価した^[7]。湯本らはヒューマノイドの手先速度の増加を評価関数とした機械学習を行い、瓦割り運動の最適化を行った(図10)。

手先速度の増加の評価関数で学習した運動を本研究で生成した評価器を用いて評価を行った。この運動の人間の主観的評価は1.0点である。この数値は目安であって教師データとしては用いないので、評価結果には反映されていない。評価結果を図11に示す。図は、教師を2~4種類としたRNNによる評価値の結果である。評価結果のデータは、それぞれ教師数や教師の評価値が異なっている。教師は1種類につき1つの主観的評価値を持っている。図の凡例は、凡例に記した数値の評価値をそれぞれ持った教師の組み合わせである。例えば凡例の一つ目は、評価値1を持つ教師と評価値0を持つ教師の組み合わせという意味である。評価結果は、それぞれの教師を組み合わせさせた教師データを基に生成した評価器を用い、生成した評価器に対象の運動の時系列データを入力して評価したものである。教師データは、人間の主観的評価が高評価と低評価であるものをそれぞれ選択した。図を見ても判る通り、中には一部評価の低い点を出した教師の組み合わせもあるが、ほとんどの教師の組み合わせで高評価を出している傾向にあり、人間の主観的評価と似ていることが判る。これは、学習によって

生成した運動フォームが整っていることが人間の主観的評価の高さに繋がったと考えられる。このことから、人間の主観的評価値と、学習によって獲得した運動の評価値が似る場合があることが判った。また、今回は手先速度の評価関数を用いたが、その他の評価関数の場合でも、RNNによる評価値が多少異なることはあるが、瓦割り運動における基本動作の違いはほとんどないため、人間の主観的評価と近い結果が出ることが考えられる。

7. 結論および今後の展望

本論文では、人間の上達過程から評価器を自動生成する手法を提案し、生成した評価器による運動の評価を行った。人間は、運動の試行回数を重ねる度にヒューマノイドの操縦を上達出来ることを確認し、それに加え、試行錯誤の回数が10回と少ない場合でも上達出来ることを確認した。また、RNNに用いる教師データがそれぞれ同じような評価値で組み合わせた場合、運動の評価の再現性が低くなってしまったことが判った。それに加え、RNNに用いる教師データがそれぞれ評価値の異なり、教師の種類が多いほど、人間の主観的評価を再現出来る傾向にあることが判った。更に、人間の主観的評価値と、学習によって獲得した運動の評価値が似る場合があることを確認した。

今後の展望では、RNNの評価が他の運動でも有効かどうか確認するためにも複数の運動で実験して比較検証する必要がある。また、今回のRNNは出力層からニューロンをフィードバックしていたが、RNNの構造を変更し、評価の精度がどのように変わるか確認する必要がある。そして、様々な評価関数で生成した運動と人間の主観的評価の関係性を調べ、運動の評価に共通する要素を割り出し、人間がどのように運動の評価をしているかを検証する。

文献

- [1] 特集「ロボットの運動学習」．日本ロボット学会誌，Vol. 22，No. 2，pp. 1-43，2004．
- [2] M. Vukobratovic and D. Juricic. Contribution to the synthesis of biped gait. *IEEE Transaction on Biomedical Engineering*, Vol. 16, No. 1, pp. 1-6, 1969.
- [3] 電気学会，GA・ニューロを用いた学習法とその応用調査専門委員会．学習とそのアルゴリズム．森北出版，2002．
- [4] 伊藤宏司．ニューロダイナミクス．共立出版，2010．
- [5] Pearlmutter and B. A. Gradient calculations for dynamic recurrent neural networks. *IEEE*, Vol. 6, No. 5, pp. 1212-1228, 1995.
- [6] Jorma Rissanen. *Stochastic Complexity in Statistical Inquiry Theory*. World Scientific Publishing Co., Inc., River Edge, NJ, USA, 1989.
- [7] Yuya Yumoto and Ikuo Mizuuchi. Exploring the possibility of mechanical energy as a multipurpose evaluation function for learning of whole body synkinetic motions. In *12th IEEE-RAS International Conference on Humanoid Robots*, pp. 632-637, 2012.