知的情報処理特論　レポート

学籍番号　T15M080
氏名　平田健人
講義日　5 月 7 日

表 1 は、データセットとそれに関連する実験パラメータをまとめている。画像特徴と時系列学習の両方に、同一の 12 層のディープニューラルネットワークが使用される。いずれの場合も、デコーダアーキテクチャは対称オートエンコーダを生じるエンコーダの鏡像である。ネットワーク構造のパラメータ設定は、経験的に、[27]及び[33]などの以前の研究に基づいて決定される。次のように 2 つのネットワークの入力と出力の次元が定義される。入力は RGB 色画素の 20×15 の行列により定義される 900 の画像特徴、出力は 10 関節角度＋30 次元の画像特徴ベクトルが 30 ステップのセグメントによって定義される 1200 の時系列学習。一次関数は両方の中央中間層で、ロジスティック関数は残りの部分で使用される。

表 1.

実験パラメータ。

	TRAIN	TEST	I / O	エンコーダ次元
IFEAT	8444	948	900	1000-500-250-150-80-30
TSEQ	20 548	776	1200	1000-500-250-150-80-30

TRAIN、TEST、I / O、及びエンコーダ次元は、それぞれ訓練データのサイズ、テストデータ、入力および出力の大きさ、及びエンコーダネットワークアーキテクチャである。

IFEAT と TSEQ はそれぞれ、画像特徴と時系列を表す

時間窓の長さは、次の 2 つの制約を考慮して決定される。最初に、時間窓の長さが増加した場合、ネットワークは、より長い文脈情報を考慮することができる。次に、時間窓の長さが長くなると、マルチモーダル時間的ベクトルの次元は、時間の許容量で処理するにはあまりにも大きくなる。暗黙的な方針は、計算制限の 3000 以下の入力次元を維持することである。マルチモーダルベクトルの次元が 40 なので、時系列の長さが 75 未満である必要がある。6 種類のオブジェクト操作行動から取得した関節角軌道の周期的な周波数を考慮すると、30 ステップは振る舞いの位相を特徴付けるために十分であることを考えた。

マルチモーダル統合学習のために、ノイズの多い値（例えば、画像特徴）を一部と、他の入力にはオリジナルの値（例えば、関節角度）を持つ例を追加した。データは三分の一ずつ画像特徴、関節角度、画像特徴＋関節角度の特徴を有する。ノイズの多い値については、元のデータの 0.1 の標準偏差でガウス雑音を重畳する。

4.3.クロスモーダルメモリ検索と時系列予測の評価

クロスモーダルメモリ検索性能を評価するために二つの実験を行った。一つ目の実験では、画像シーケンスを使って関節角度シーケンス（動作）を生成し、一方、関節角度シーケンスを使って画像シーケンスを生成した。これらの実験では、完全な 30 ステップのいずれかのモダリティへの入力が提供され、他のモダリティのシーケンスが生成された。時系列予測を評価する実験では、入力の時間窓の長さはT_{in}=25 として定義され、対応する将来の 5 つのステップは内部予測で生成される。以上の実験のすべての設定、周期入力の初期値はランダムに生成されているが、内部値は最終的にネットワークの汎化能力によって他のモダリティの入力値に関連して対応する状態に収束する。

図 5 は、画像シーケンスの入力と時系列予測から関節角度シーケンス生成の結果の例を示している。反復検索された関節角度ベクトルを累積して、オブジェクト操作行動の完全な長さの軌跡を生成した。同図において、最上段のグラフ(図 5 の(a))は、正しい元データである。第二列(図 5 の(b))上のグラフは、完全な画像シーケンスと時系列予測をクロスモーダルメモリ検索により再構成された関節角度。一番下の行のグラフ(図 5(c))はマルチモーダル時系列の最後の 5 ステップで、25 ステップの関節角度を予測することを示している。最初の 30 ステップの低い復興品質が生成プロセスの初期の反復で周期入力に指定されたランダムな値に起因している。

図 6 は、関節角度シーケンス入力から画像シーケンスの生成の結果の例を示している。図に示されている画像は、それぞれの行動のための一連の画像から引き出された一つの画像である。画像の内容が多少異なるものの、画像中に現れたオブジェクトは正しく再構築され、カラーブロブの位置が適切に運動の位相に同期される。

ネットワークのための 10 種類の初期モデルパラメータの設定を準備し、6 種類のオブジェクト操作行動から全く同じ初期値とデータで実験することによって、クロスモーダルメモリ検索の定量的評価を行った。表 2 は、これらの結果をまとめたものである。表では、IMG→MTN は運動から画像を、MTN→IMG は、画像から運動を示す。さらに、6 種類の行動パターンから時系列予測（PRED）性能も示されている。表の各エントリに与えられた数値は、再構成された軌跡エラーの平方根を意味する。(0 と 1 の間に正規化)表 2 のエラーの平方根は、再構成エラーが評価条件の全てについて 10%以下であることを示す。

表 2.

再生誤差.

	IMG → MTN	MTN → IMG	PRED
LIFT	7.11×10^{-2} (1.44×10^{-3})	1.76×10^{-2} (8.99×10^{-4})	3.91×10^{-2} (6.47×10^{-4})
ROLL	7.05×10^{-2} ($1.55\times$	4.45×10^{-2} ($1.20\times$	4.41×10^{-2} ($7.33\times$

	IMG → MTN	MTN → IMG	PRED
	10 -3)	10 -3)	10 -4)
RING-L *	4.95×10 -2 (2.64× 10 -3)	1.83×10 -2 (4.72× 10 -4)	2.21×10 -2 (8.19× 10 -4)
RING-R *	3.64×10 -2 (2.61× 10 -3)	1.79×10 -2 (3.64× 10 -3)	1.98×10 -2 (4.90× 10 -4)
PLT *	8.98×10 -2 (1.35× 10 -3)	1.49×10 -2 (2.96× 10 -3)	3.94×10 -2 (4.34× 10 -4)
RWY *	5.63×10 -2 (9.50× 10 -4)	1.89×10 -2 (5.32× 10 -3)	2.75×10 -2 (4.32× 10 -4)

*

LIFT、ROLL、RING-L、RING-R、PLT、及び RWY はそれぞれ、ボールリフト、ボールロール、ベルリング L、ベルリング R、プレート上のボールロール、およびロープウェイを表す

**カッコ内の標準偏差。

エラーの平方根の各々は、以下のように計算される。

$$E_{IMG\rightarrow MTN} = \sqrt{\frac{1}{T_{seq}}\sum_{t=1}^{T_{seq}} |\tilde{a}_t - \hat{a}_t|^2}, \tag{12}$$

$$E_{MTN\rightarrow IMG} = \sqrt{\frac{1}{T_{seq}}\sum_{t=1}^{T_{seq}} |\tilde{r}_t^i - \hat{r}_t^i|^2}, \tag{13}$$

$$E_{PRED} = \sqrt{\frac{1}{T_{seq}}\sum_{t=1}^{T_{seq}} |\tilde{s}_t - \hat{s}_t|^2}, \tag{14}$$

$E_{IMG\rightarrow MTN}$、$E_{MTN\rightarrow IMG}$、E_{PRED}は、添字によって識別再構築モードに対応するエラーの平方根。 $\tilde{a}_t$、 $\hat{a}_t$、 $\tilde{r}_t^i$、 $\hat{r}_t^i$、 $\tilde{s}_t$、 $\hat{s}_t$は、生画像データ、関節角度、時間 t におけるマルチモーダル特徴を表す真理と再構成されたベクトルである。そして、T_{seq}はテストの繰り返しの長さである。

最後に、より詳細に時系列の予測性能を分析するために、時間窓の長さのステップT_{in}を 25～5 に小さく変化させることにより、時間窓の最後のステップで予測誤差を評価する。エラーの平方根が予想されるように、予測長が増加すると、その予測誤差が大きくなることを示している。それにもかかわらず、再構成エラーは、評価条件のすべてで 10%未満である。

補足資料

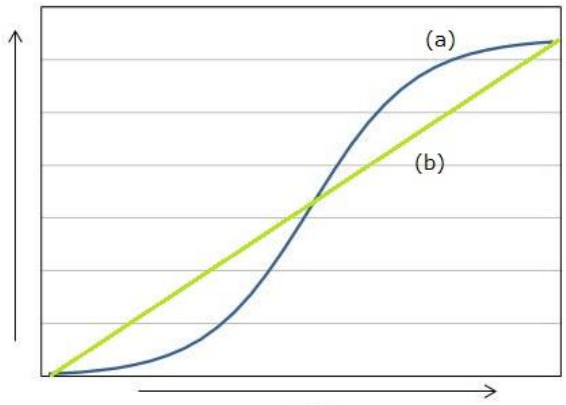
図 3 : マルチモーダル行動学習と取得機構 ; 二つの独立したディープニューラルネットワークは、(a)画像圧縮と(b)時系列学習のために利用される。

図 4 : オブジェクト操作行動。(a)ボールリフト : その後、上下動きと肩の高さに 3 回ボールを上げ、両手でテーブルの上に黄色のボールを保持している。(b)ボールロール : 交互に腕の動きを使用してテーブルの上の青いボールを左右に反復して転がす。(c)と(d)ベルリング L / R : 対応するアームの運動によって、テーブルの右または左のどちらにも配置された緑の鐘を鳴らす。(e) プレート上のボールロール : 両手に取り付けたプレートにオレンジ色のボール入れ、転がすと交互に上下両腕を振る。(f)ロープウェイ : 交互に上下両腕を動かすことで、両手に装着したひもにぶら下がっている赤いボールをスイングする。

図 5 : 提案モデルによる例モーション再構成 : 最上段のグラフ(a)は、試験データの元の運動軌跡を示す。第 2 の行(b)及び下段(c)のグラフは、それぞれ、画像シーケンスと時系列予測からのクロスモーダルメモリ検索により取得された再構成された軌跡を示す。再構築された軌道は、一番上の行に示されたのと同じ行動に対応している。

図 6 : 提案モデルによる例画像再構成 ; 一番上の行の画像(a)は、試験データにおける画像特徴ベクトルから伸張原画像を示す。下段の画像(b)は、関節角度配列からクロスモーダルメモリ検索により取得された特徴ベクトルから対応する再構成された伸張画像を示す。

図 7 : 予測の長さに応じて 6 オブジェクト操作行動の時間的シーケンス予測誤差; 平均および標準偏差が 10 レプリケート学習実験から計算される。プロットは、水平誤差棒の重複を避けるために、元の位置からずれている。



(a):ロジスティック関数

(b):一次関数