AI 特許紹介(19)
AI 特許を学ぶ！究める！
～Deep LSTMP～

2020 年 8 月 7 日
河野特許事務所
所長　弁理士　河野英仁

「AI 特許紹介」シリーズは、注目すべき AI 特許のポイントを紹介します。熾烈な競争となっている第４次産業革命下では AI 技術がキーとなり、この AI 技術・ソリューションを特許として適切に権利化しておくことが重要であることは言うまでもありません。

AI 技術は Google, Microsoft, Amazon を始めとした IT プラットフォーマ、研究機関及び大学から毎週のように新たな手法が提案されており、また AI 技術を活用した新たなソリューションも次々とリリースされています。

本稿では米国先進 IT 企業を中心に、これらの企業から出願された AI 特許に記載された AI テクノロジー・ソリューションのポイントをわかりやすく解説致します。

1.概要
特許権者　Google
出願日　2017 年 3 月 9 日
登録日　2018 年 7 月 17 日
登録番号 US10026397
発明の名称　リカレントプロジェクション層を含む長期短期記憶（LSTM）ニューラルネットワークを使用した音響シーケンスの処理

397 特許は、リカレントニューラルネットワークの一種である LSTM にリカレントプロジェクション層を組み合わせることにより、モデルのメモリサイズ及び汎化性能の双方を向上させるアイデアである。

2.特許内容の説明
下記図１は音響モデリングシステムの構成を示すブロック図である。

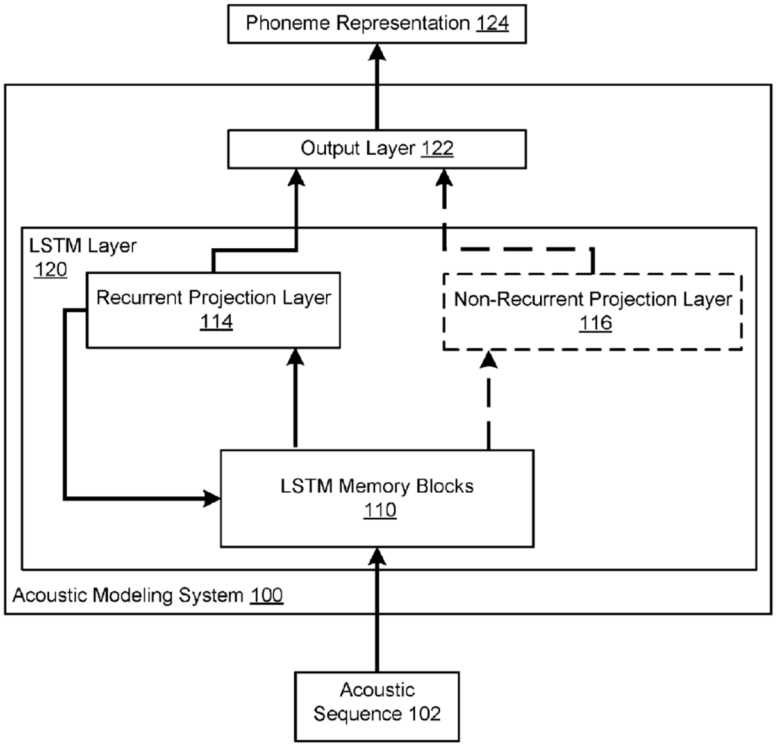

図 1

音響モデリングシステム 100 は、1 つまたは複数の LSTM 層 120、及び、出力層 122 を含む。図 1 では簡略化のため 1 つの LSTM 層を記載しているが、実装の際には最低 LSTM 層から最高 LSTM 層までの複数層が用いられる。

各タイムステップで、各 LSTM 層は前の LSTM 層からの入力を受け取る。また、LSTM 層が LSTM 層のシーケンスの最下位層である場合は、タイムステップの音響特徴表現を受け取り、当該タイムステップで層出力を生成する。

各 LSTM 層は、LSTM メモリブロック 110 およびリカレントプロジェクション層 114 を含む。LSTM メモリブロック 110 は、LSTM 層 120 によって受信された入力、例えば、現在のタイムステップの音響表現または先行する LSTM 層によって生成された層出力を処理して、タイムステップの LSTM 出力を集合的に生成する。

リカレントプロジェクション層 114 は、LSTM メモリブロック 110 によって生成された LSTM 出力を受け取り、リカレントプロジェクション層のパラメータのセットの現在の値に従って、LSTM 出力からリカレントプロジェクション出力を生成する。

一般に、リカレントプロジェクション層 114 は、リカレントプロジェクション層のパ

ラメータの現在の値に従って、LSTM 出力を低次元空間に投影する。すなわち、リカレントプロジェクション出力は、リカレントプロジェクション層 114 によって受信される LSTM 出力よりも低い次元性を有する。例えば、リカレントプロジェクション層 114 によって受信された LSTM 出力は、リカレントプロジェクション出力の次元性の約 2 倍である次元性を有する。例えば 1000:500、または、2000:1000 である。

音響モデリングシステム 100 は、所定のタイムステップでリカレントプロジェクション層 114 によって生成されたリカレントプロジェクション出力を、LSTM メモリブロック 110 に提供して、音響シーケンスの次のタイムステップの LSTM 出力を生成する際に使用する。

LSTM 層 120 は、オプションで、非リカレントプロジェクション層 116 をも含む。非リカレントプロジェクション層 116 は、LSTM メモリブロック 110 によって生成された LSTM 出力を受け取り、LSTM 出力に対する一組のパラメータの現在の値に従って LSTM 出力から非リカレントプロジェクション出力を生成する。

非リカレントプロジェクション層 116 は、LSTM 出力をリカレントプロジェクション層 114 と同じく低次元空間に投影するが、リカレントプロジェクション層 114 とは、異なるパラメータ値を使用する。

LSTM 層 120 が非リカレントプロジェクション層 116 を含む場合、音響モデリングシステム 100 は、リカレントプロジェクション出力および非リカレントプロジェクション出力を、LSTM 層 120 の層出力として提供する。

LSTM 層 120 が非リカレントプロジェクション層 116 を含まない場合、音響モデリングシステム 100 は、LSTM 層 120 の層出力としてリカレントプロジェクション層 114 の出力を提供する。

出力層 122 は、LSTM 層のシーケンスの最も高い LSTM 層から層出力を受け取り、出力層のパラメータのセットの現在の値に従って、現在のタイムステップのスコアのセットを生成する。シーケンス内の各タイムステップのスコアのセットが生成されると、音響モデリングシステム 100 は、各タイムステップのスコアのセットを含む音素表現を出力する。

音響モデリングシステム 100 は、各タイムステップで最高のスコアを有する音素または音素サブディビジョンを選択し、選択した音素または音素サブディビジョンのシー

ケンスを音響シーケンスの音素表現として出力することができる。

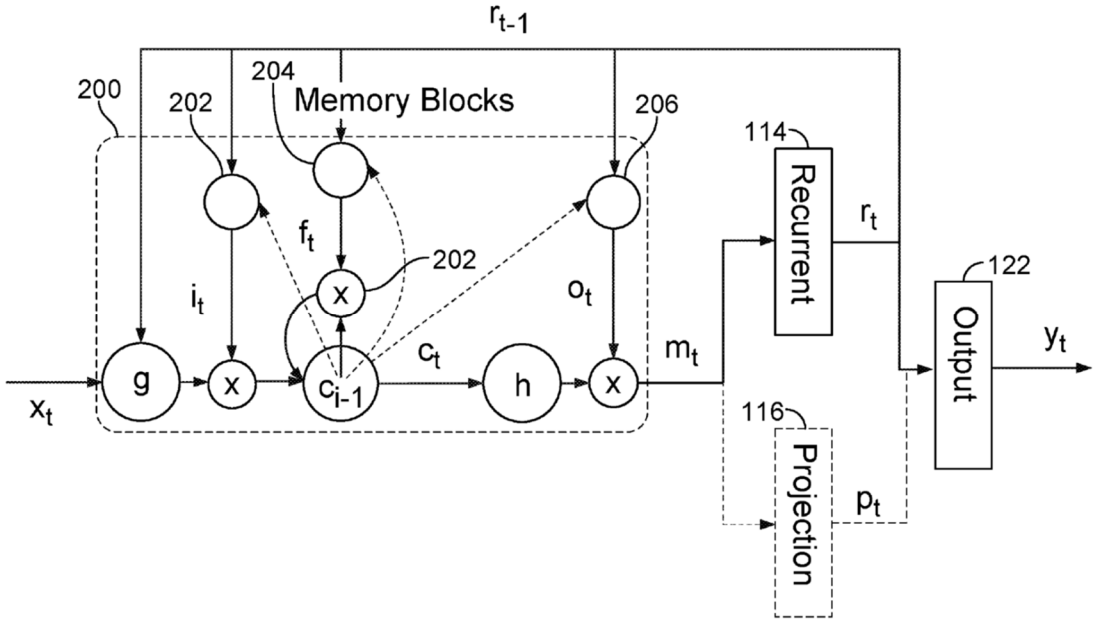

図２

図２は、LSTM メモリブロック 200 の例を示す。LSTM メモリブロック 200 は、入力 x_t を受信し、入力から、および以前のリカレントプロジェクション出力 r_{t-1} から出力 m_t を生成する LSTM メモリセルを含む。

LSTM メモリセルは、メモリセルへの入力アクティベーションのフローを制御する入力ゲート 202、セルの出力フローを制御する出力ゲート 204、および、セルの状態を介して、セルへの入力として内部状態を追加する前に、セルの内部状態をスケーリングする忘却ゲート 206 を含む。

セルは出力 m_t を計算し、m_t が次の方程式を満たすようにする。

$$i_t=\sigma(W_{ix}x_t+W_{ir}r_{t-1}+W_{ic}c_{t-1}+b_i)$$
$$f_t=\sigma(W_{fx}x_t+W_{rf}r_{t-1}+W_{cf}c_{t-1}+b_f)$$
$$c_t=f_t\odot c_{t-1}+i_i\odot g(w_{cx}x_i+W_{cr}r_{t-1}+b_c)$$
$$o_t=\sigma(W_{ox}x_t+W_{or}r_{t-1}+W_{oc}c_t+b_o)$$
$$m_i=o_t\odot h(c_i)$$

ここで、i_t は現在のタイムステップでの入力ゲートのアクティベーション、f_t は現在のタイムステップでの忘却ゲートのアクティベーション、o_t は現在のタイムステップでの出力ゲートのアクティベーション、c_t は現在のタイムステップでのセルのアクティ

ベーション、c_{t-1}は、前のタイムステップでのセルのアクティベーションである。

⊙は要素ごとの積演算、gはセル入力アクティベーション関数、hはセル出力アクティベーション関数、各W項はLSTMメモリセルの現在の重み値のそれぞれの行列、b_i, b_f, b_c,及びb_oはバイアスベクトルである。

出力m_tが計算されると、リカレントプロジェクション層114は、出力m_t使用して現在のタイムステップに対するリカレントプロジェクション出力r_tを計算する。リカレントプロジェクション出力rtは次の条件を満たす。

$$r_t = W_{rm} m_t$$

ここで、W_{rm}は、リカレントプロジェクション層114の重みの現在値の行列である。次に、リカレントプロジェクション出力r_tは、音素表現の計算に使用するために出力層122に提供されるか、または、シーケンスの次のLSTM層に提供され、音響シーケンスの次のタイムステップで、出力m_{t+1}の計算に使用するためにメモリセルにフィードバックされる。

非リカレントプロジェクション層116が含まれる場合、非リカレントプロジェクション層116は、出力m_tを使用して現在のタイムステップに対する非リカレントプロジェクション出力p_tを計算する。リカレントプロジェクション出力p_tは次の条件を満たす。

$$p_t = W_{pm} m_t$$

ここで、W_{pm}は、非リカレントプロジェクション層116の重みの現在値の行列である。非リカレントプロジェクション出力p_tは、リカレントプロジェクション出力r_tと組み合わせて、音素表現の計算に使用するために出力層122に提供され、あるいは、シーケンス内の次のLSTM層に提供されるが、メモリセルにはフィードバックされない。

出力層122は、リカレントプロジェクション出力r_tと、オプションで、最も高いLSTM層によって生成された非リカレントプロジェクション出力p_tとを受け取り、現在のタイムステップのスコアベクトルy_tを計算する。

現在のタイムステップのスコアベクトルには、HMM(Hidden Markov Model)状態のセットのそれぞれのスコアが含まれる。例えば、それぞれのスコアは、HMM状態のセットにわたる確率分布を定義する確率である。出力層122が非リカレントプロジェクション出力を受け取っている場合、スコアベクトルy_tは以下を満たす。

$$y_t = W_{yr} r_t + W_{yp} p_t + b_y$$

出力層 122 が非リカレントプロジェクション出力を受け取っていない場合、スコアベクトル y_t は以下を満たす。

$$y_t = W_{yr} r_t + b_y$$

ここで、b_y は出力層 122 のバイアスベクトルであり、各 W 項は出力層 122 の重みの現在値のそれぞれの行列である。

3.クレーム

397 特許のクレーム 1 及び 2 は以下の通りである。

1. 音響特徴表現を処理するためのシステムにおいて、

1 つまたは複数の長期短期記憶（LSTM）層で、それぞれがリカレントプロジェクション層を含み、1 つまたは複数の LSTM 層は、最低の LSTM 層から最高の LSTM 層までの順序で配置され、1 つまたは複数の LSTM 層のそれぞれは、以下を含む動作を実行するように構成される：

音響特徴表現またはシーケンス内の先行する LSTM 層により生成された層出力である層入力を受信し、

層入力と以前のリカレントプロジェクション出力に基づいて LSTM 出力を生成し、

それぞれのリカレントプロジェクション層を処理することにより、LSTM 出力を低次元空間に投影すべく、重みの現在値の行列を適用してリカレントプロジェクション出力を生成し、

以前のリカレントプロジェクション出力をリカレントプロジェクション出力で更新し、該更新された以前のリカレントプロジェクション出力は、次の LSTM 出力を生成する際にシーケンスの次の LSTM 層によって使用され、

以下を含む第 2 オペレーションを実行するように構成された出力層を備え：

シーケンスの最上位の LSTM 層によって生成されたリカレントプロジェクション出力に従って決定されたスコアのセットを生成し、

該スコアのセットは、それぞれの音素または音素サブディビジョンが音響特徴表現を表す可能性を表す複数の音素または音素サブディビジョンのそれぞれのスコアを含む。

2.クレーム 1 のシステムにおいて、

LSTM 層のそれぞれは、非リカレントプロジェクション層を含み、

LSTM 出力は、以前の非リカレントプロジェクション出力にさらに基づいており、前

記オペレーションはさらに以下が含まれる：

それぞれの非リカレントプロジェクション層を介して処理することにより、LSTM 出力を低次元空間に投影すべく、重みの現在値の他の行列を適用して非リカレントプロジェクションを生成し、

以前の非リカレントプロジェクション出力を非リカレントプロジェクション出力で更新し、該更新された以前の非リカレントプロジェクション出力は、次の LSTM 出力を生成する際にシーケンスの次の LSTM 層によって使用される。

4. DLSTMP に関する論文

リカレントプロジェクションを用いた深層 LSTM に関する論文[1]が Google の Hasim Sak 氏らにより発表されている。

下記図 2 は LSTM RNN のアーキテクチャを示す説明図である。

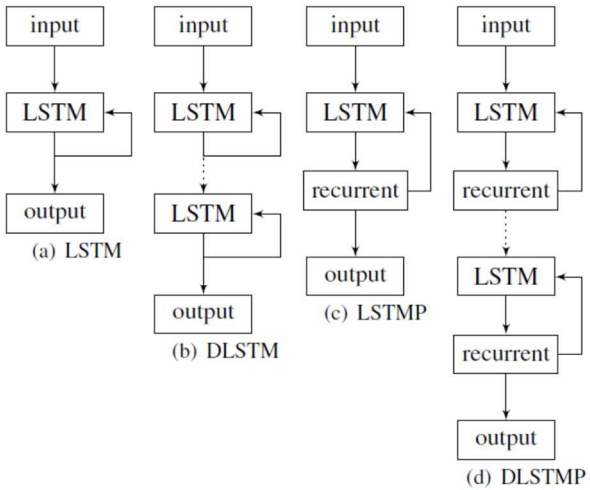

Figure 2: LSTM RNN architectures.

自然言語処理及び時系列データの処理には LSTM(図 2(a))を深層化したディープ LSTM(DLSTM 図 2(b))が用いられる。

本論文では、DLSTM と同様に、それぞれが個別のリカレントプロジェクション層を持つ複数の LSTM レイヤーがスタックされた DLSTMP(図 2(c))が提案されている。

---

[1] Has¸im Sak, Andrew Senior, Franc¸oise Beaufays, “LONG SHORT-TERM MEMORY BASED RECURRENT NEURAL NETWORK ARCHITECTURES FOR LARGE VOCABULARY SPEECH RECOGNITION” arXiv:1402.1128v1 2014 年 2 月 5 日

LSTMP を使用すると、出力層及びリカレント接続とは独立してモデルのメモリを増やすことができる。

ただし、メモリサイズを増やすと、入力シーケンスデータを記憶することにより、モデルが過剰適合しやすくなる。DNN は、深さが増すにつれ、未知の例に対しより一般化できることがわかっている。深さは、訓練データに対し、モデルをより過剰適合しにくくする。これは、ネットワークへの入力は、多くの非線形関数を通過する必要があるためである。

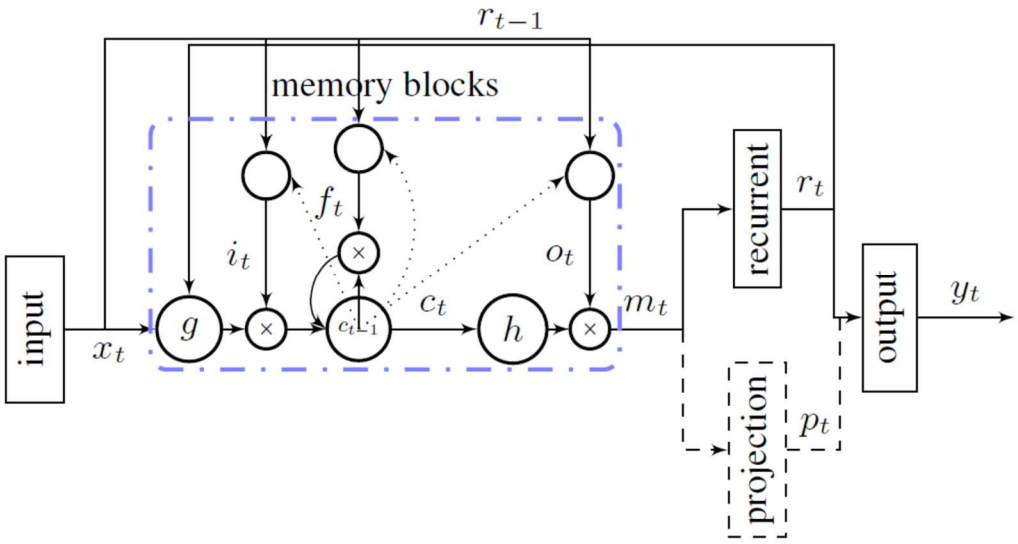

図 1　リカレントプロジェクション層及び非リカレントプロジェクション層を含む DLSTM ネットワーク（簡略化のため 1 メモリブロックのみを示している）

この動機から、この論文では、モデルのメモリサイズと汎化性能を向上させることを目的としたディープ LSTMP(DLSTMP 図 2(d)、図 1)アーキテクチャについて実験している。

実験結果を下記図 3 に示す。

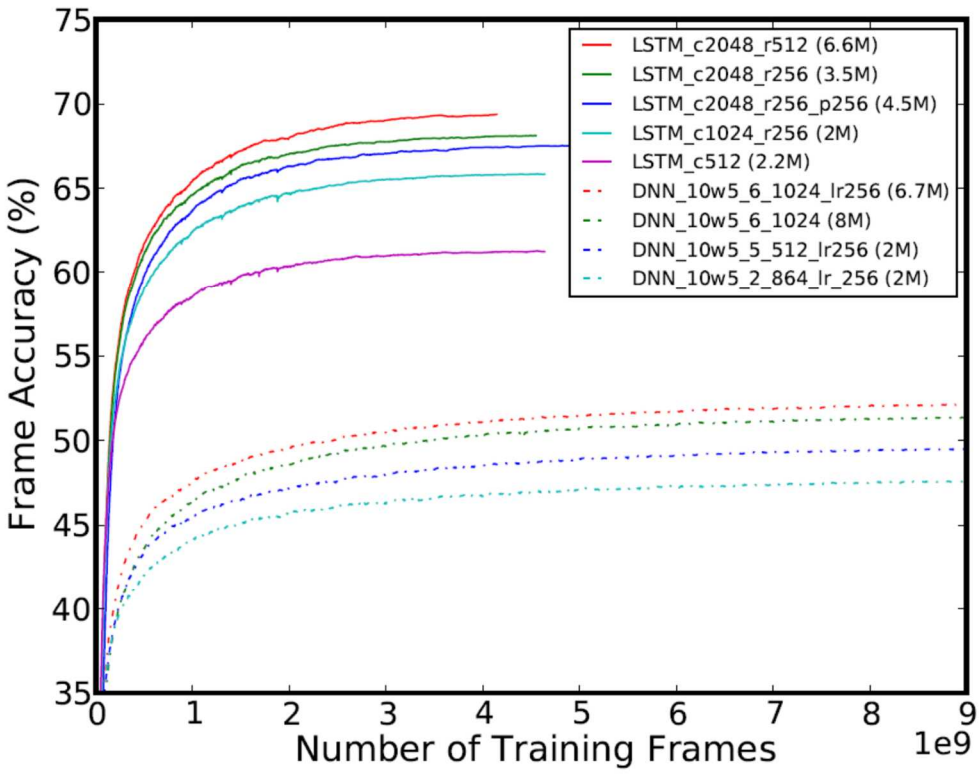

Fig. 3. 2000 context dependent phone HMM states.

図 3 のグラフ系列において、 cN は、LSTM のメモリセルの数（N）と RNN の隠れ層のユニット数を示す。 rN は、LSTM および RNN のリカレントプロジェクションユニット数を示す。 pN は、LSTM の非リカレントプロジェクションユニット数を示す。各モデルのパラメータの数は括弧内に示されている。

図 3 に示すように、LSTM ネットワークは、DNN よりもはるかに優れたフレーム精度を提供しながら、収束を高速化している。

また図 3 の LSTM512(紫)と、LSTM1024_r256(水色)と比較すればわかるように、提案されている LSTM プロジェクション RNN アーキテクチャは、同じ数のパラメータを持つ標準の LSTM RNN アーキテクチャよりもはるかに優れた精度を有している。

以上

著者紹介
河野英仁
河野特許事務所、所長弁理士。立命館大学情報システム学博士前期課程修了、米国フラ

ンクリンピアースローセンター知的財産権法修士修了、中国清華大学法学院知的財産夏季セミナー修了、MIT(マサチューセッツ工科大学)コンピュータ科学・AI 研究所 AI コース修了。

AI 特許コンサルティング、医療 AI 特許コンサルティングの他、米国・中国特許の権利化・侵害訴訟を専門としている。著書に「世界のソフトウエア特許(共著)」、「FinTech 特許入門」、「AI/IoT 特許入門 2.0」がある。