AI 特許紹介(13)

～ニューラルトランスデューサ～

2020 年 4 月 10 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　2016 年 11 月 11 日

登録日　2018 年 9 月 25 日

登録番号 US10043512

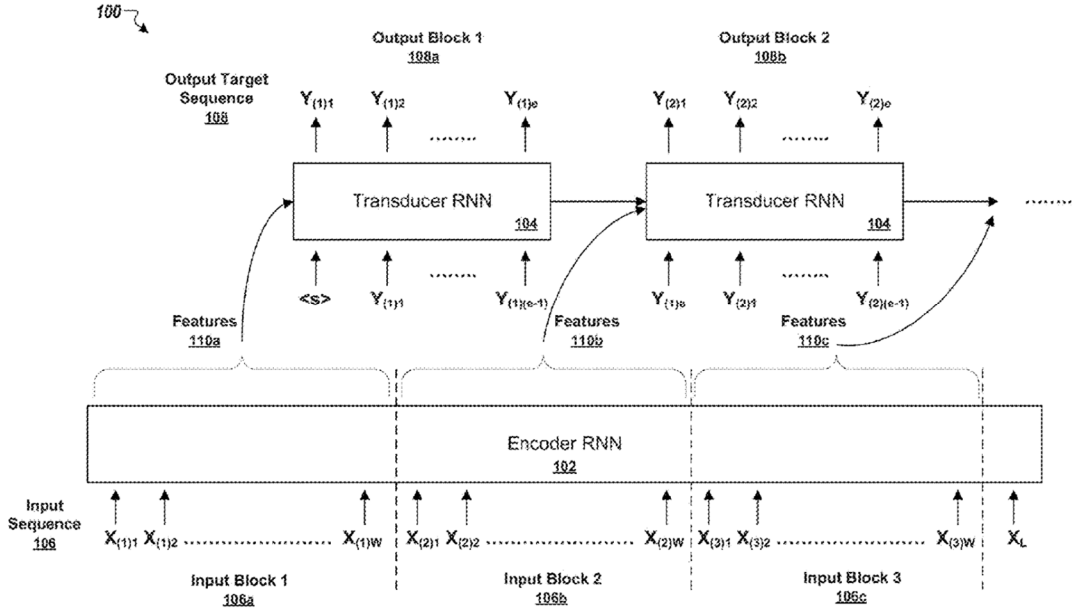
発明の名称　部分条件付けを使用して入力シーケンスからターゲットシーケンスを生成する

512 特許は、エンコーダ RNN（リカレントニューラルネットワーク）とトランスデューサ RNN を使用して、音声入力途中の段階でもリアルタイムで翻訳または言語予測を行うアイデアに関する。

2.特許内容の説明

リカレントニューラルネットワークは、入力シーケンスを受け取り、入力シーケンスから出力シーケンスを生成するニューラルネットワークである。一連のセンテンスが RNN に入力された場合、多言語に変換または次に現れるセンテンスが予測される。

しかしながら、センテンス全体の入力を待っていればリアルタイムでの処理に適さないという問題がある。512 特許ではエンコーダ RNN とトランスデューサ RNN とを組み合わせ、区切られたセンテンスを順次リアルタイムで処理することとしている。

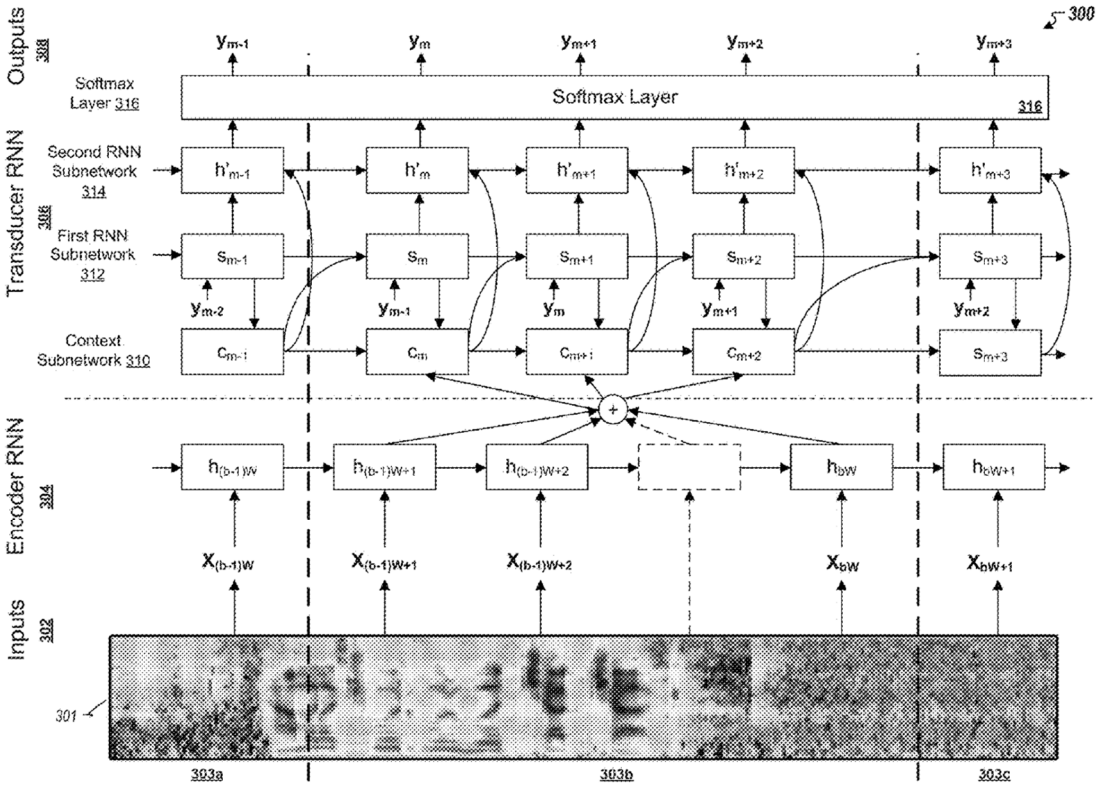


具体的には、入力シーケンスの固定数の入力タイムステップで構成される各ブロックについて、システムはエンコーダーリカレントニューラルネットワーク（RNN）102 を使用して入力タイムステップのブロック 106a,106b,106c の各入力を処理し、入力のそれぞれの特徴表現 110a,110b.110c を生成する。RNN には LSTM(Long short-term memory)を使用する。

次に、システムは、トランスデューサ RNN104 を使用して、(i) ブロック内の入力の特徴表現 110、および（ii）現在の出力タイムステップの直前の先行出力タイムステップでの先行出力を処理する。そしてトランスデューサ RNN104 は、直前の出力タイムステップの直後の 1 つ以上の出力タイムステップのそれぞれの出力を選択する。

このように、トランスデューサ RNN104 を使用することで、エンコーダ RNN102 が入力シーケンスのすべての入力の特徴表現を生成する前に、ターゲットシーケンス 108 の出力の選択を開始することができる。

なお、519 特許にはより拡張したシステムが以下の通り開示されている。



3.クレーム

512 特許のクレーム 1 は以下の通りである。

1. 複数の入力タイムステップのそれぞれでの各入力を含む入力シーケンスから、複数の出力タイムステップのそれぞれでの各出力を含むターゲットシーケンスを生成する方法において、

入力シーケンスの固定数の入力タイムステップの各ブロックに対して、

入力の各特徴表現を生成するために、エンコーダリカレントニューラルネットワーク（RNN）を使用して、入力タイムステップのブロック内の各入力を処理し、

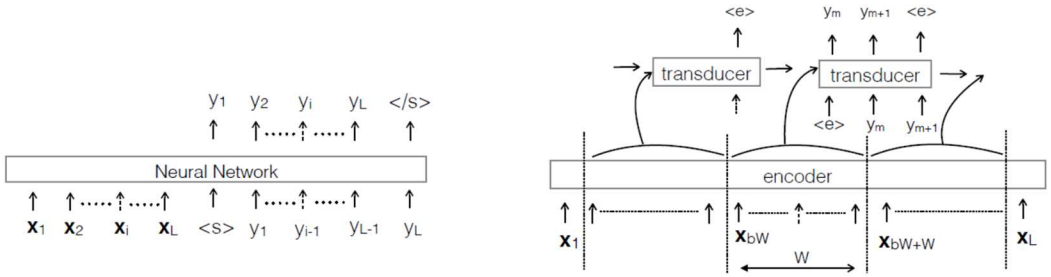
ブロックに対応する複数のタイムステップの一部の出力を選択し、これには、複数のタイムステップの一部の各現在の出力タイムステップが含まれ、

トランスデューサ RNN を使用して、現在の出力タイムステップに対応する出力を選択するために、(i) ブロック内の入力の特徴表現に基づくデータ、および (ii) 現在の出力タイムステップの直前の先行出力タイムステップでの先行出力を処理し、

現在の出力タイムステップの各出力が指定されたブロックの終わりの出力である場合、ブロックのそれ以上の出力の生成を控える。

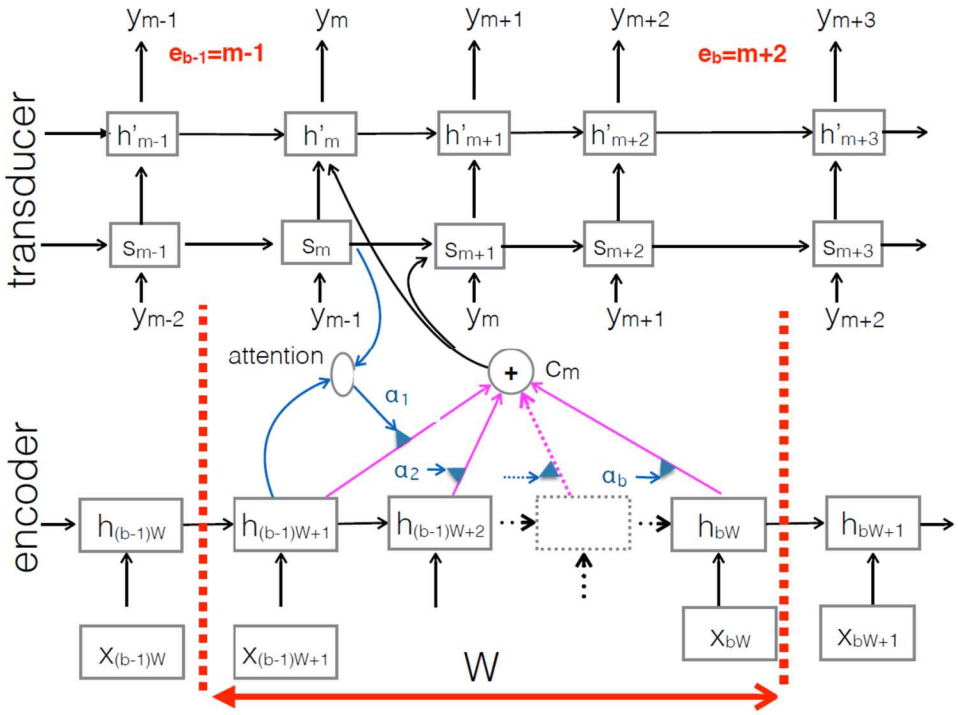
4. ニューラルトランスデューサの論文

ニューラルトランスデューサに関する論文[1]が、Google の Navdeep Jaitly 氏らにより発表されている。



(a) seq2seq　　(b) Neural Transducer

上図の(a)は一般的な seq2seq モデルを示している。時系列のデータ x_1-x_L が入力され、予測される時系列データ y_1-y_L が出力される。これに対し、論文におけるニューラルトランスデューサでは、エンコーダによりブロック毎に入力データの特徴量がトランスデューサに入力され、トランスデューサから予測される時系列データがリアルタイムで出力される。



上図は論文に示されているニューラルトランスデューサアーキテクチャである。エンコーダは、入力された音声データ x_i（i=1・・L）から、各タイムステップ i において、隠れ状態ベクトル h_i を生成する。

トランスデューサは、各ステップで入力ブロックを取得し、seq2seq モデルを使用し

[1] Navdeep Jaitly “A Neural Transducer” 2016 年 8 月 4 日

て最大 M の出力トークンを生成する。トランスデューサーは、前の出力タイムステップへのリカレントコネクションを使用して、ブロック全体でその状態を維持する。上図では、ブロック b のトークンを生成するトランスデューサが示されており、このブロック b で最終的に出力されるサブシーケンスは $y_m y_{m+1} y_{m+2}$ である。

下記表は 1 ブロックを 15 フレームとしたニューラルトランスデューサと、通常の seq2seq モデルとを比較した実験結果である。

W	BLOCK-RECURRENCE	PER
15	No	34.3
15	Yes	20.6

ニューラルトランスデューサのほうが PER(Phone Error Rate)が 20.6 と大幅に低減している。論文ではブロックサイズを適宜変更することでさらなる精度向上が図れる点、述べている。

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。

以上