AI 特許紹介(14)

～ニューラル画像キャプションジェネレータ～

2020 年 5 月 8 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　2015 年 11 月 13 日

登録日　2018 年 1 月 2 日

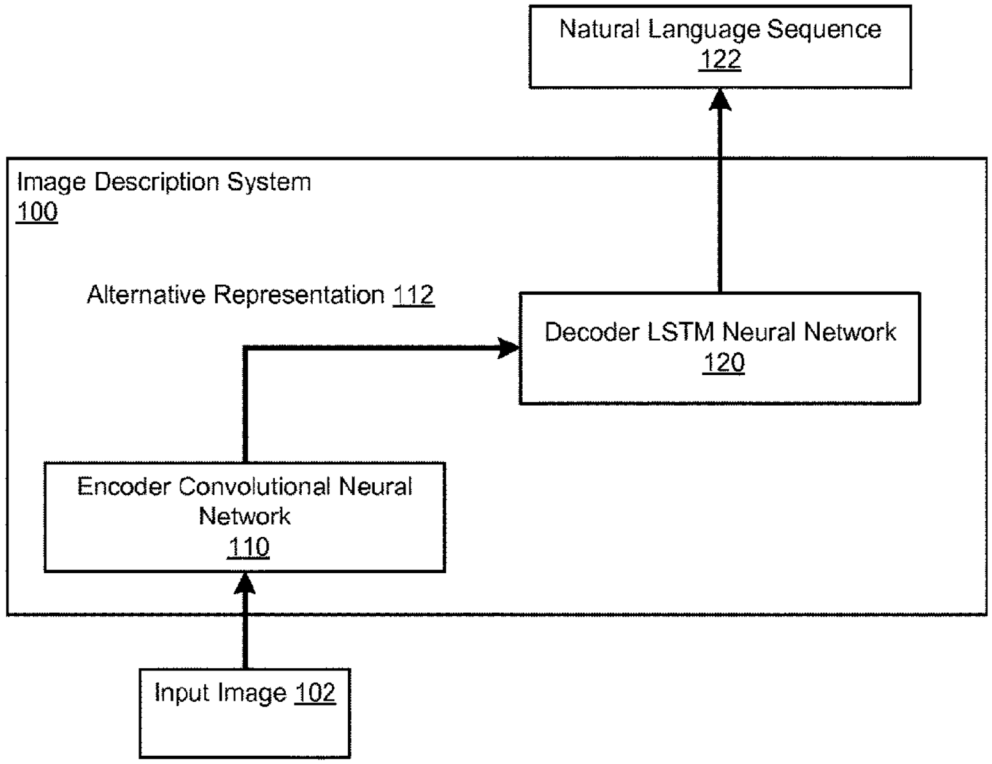
登録番号 US9858524

発明の名称　画像の自然言語記述の生成

524 特許は、入力画像から当該画像のキャプションを生成するニューラルネットワークに関する。

2.特許内容の説明

下記図は画像記述システム 100 の構成を示すブロック図である。



画像記述システム 100 は、入力画像 102 を処理するエンコーダ畳み込みニューラルネットワーク 110、及び、エンコーダ畳み込みニューラルネットワーク 110 から出力された画像の代替表現 112 を処理して自然言語シーケンス 122 を出力するデコーダ LSTM ニューラルネットワーク 120 により構成される。

エンコーダ畳み込みニューラルネットワーク 110 は、入力画像 102 を受け取り、パラメータのセットに従って入力画像 102 から代替表現 112 を生成する畳み込みニューラルネットワークである。エンコーダ畳み込みニューラルネットワーク 110 は例えば画像の特徴を抽出するコアニューラルネットワークと、カテゴリ分類を行うソフトマックス層等の出力層とを含むニューラルネットワークから、後者の出力層を除いたものである。代替表現 112 は、エンコーダ畳み込みニューラルネットワーク 110 の最後のコアレイヤーの出力である。

デコーダ LSTM ニューラルネットワーク 120 は、1 つ以上の LSTM ニューラルネットワーク層を含む LSTM ニューラルネットワークであり、各 LSTM 層は、1 つ以上の LSTM メモリブロックを含む。各 LSTM メモリブロックには、1 つ以上のセルを含み、各セルには、入力ゲート、忘却ゲート、および出力ゲートが含まれる。

デコーダ LSTM ニューラルネットワーク 120 はエンコーダ畳み込みニューラルネッ

トワーク 110 から出力された代替表現を処理し、入力画像の自然言語シーケンスを生成する。自然言語シーケンスとは、出力順序に従って配置された、ターゲット自然言語の一連の単語である。デコーダーLSTM ニューラルネットワーク 120 とエンコーダ畳み込みニューラルネットワーク 110 は、対応する入力画像の記述である自然言語シーケンスを生成するようにトレーニングされている。

システムは、可能性のある自然言語シーケンスごとに、各候補自然言語シーケンスのそれぞれのシーケンススコアを生成する。システムは、最高のシーケンススコアを持つ自然言語シーケンスを入力画像の自然言語シーケンスとして選択する。

3.クレーム

524 特許のクレーム１6 は以下の通りである。

16. 1 つまたは複数の非一時的なコンピューター記憶媒体にエンコードされたコンピュータープログラム製品において、該コンピュータープログラム製品は、1 つまたは複数のコンピューターにより実行された場合に、1 つまたは複数のコンピューターに下記命令を実行させる：

入力画像を取得し、

入力画像の代替表現を生成すべく、第 1 のニューラルネットワークを使用して入力画像を処理し、

入力画像を記述するターゲット自然言語で複数の単語のシーケンスを生成するために、長期短期記憶（LSTM）ニューラルネットワークを使用して入力画像の代替表現を処理し、

シーケンス内の単語は出力順序に従って配置され、

入力画像の代替表現を処理することは、出力順序の初期位置について、さらに以下を含む：

(i) 単語のセット内の各単語のそれぞれの初期単語スコアを生成するために、LSTM ニューラルネットワークを使用して特別な開始単語を処理し、

(ii) 初期単語スコアを使用して、出力セットの初期位置にある単語として、単語のセットから単語を選択する。

4. ニューラル画像キャプションジェネレータの論文

ニューラル画像キャプション(NIC)ジェネレータに関する論文が Oriol Vinyals 氏らにより発表されている[1]。

[1] Oriol Vinyals, Alexander Toshev, Samy Bengio, Dumitru Erhan: A Neural Image Caption Generator arXiv:1411.4555v2

論文には画像 CNN とこの画像 CNN に続く言語生成 RNN を組み合わせ、下記図 1 に示すように画像を入力することで文章を生成するモデルが詳述されている。

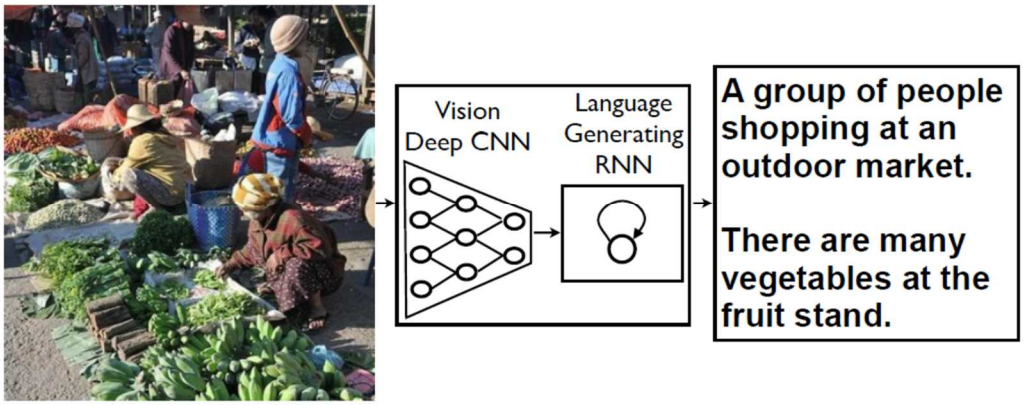


図 1

RNN を用いた機械翻訳が実現できているが、同様のコンセプトにより画像の特徴量を抽出し当該画像を文章で表現せんとするものである。モデルは、下記式で示す適切な記述の可能性を最大にする必要がある。

$$\theta^{\star} = \arg\max_{\theta} \sum_{(I,S)} \log p(S|I;\theta) \qquad (1)$$

ここで、θ はモデルのパラメータ、I は画像、S は正しい文章である。また θ の依存性を削除すると下記式で表現できる。S の長さは一定でないため、同時確率をモデリングするために $S_0, \ldots, S_N$ に対し（N は特定の例の長さ)チェーンルールを適用する。

$$\log p(S|I) = \sum_{t=0}^{N} \log p(S_t|I, S_0, \ldots, S_{t-1}) \qquad (2)$$

トレーニングの際には、(S,I) がトレーニングのペアであり、式 2 で示されるログの確率の総和を、全てのトレーニングセットに対して確率的勾配効果を用いて、最適化する。

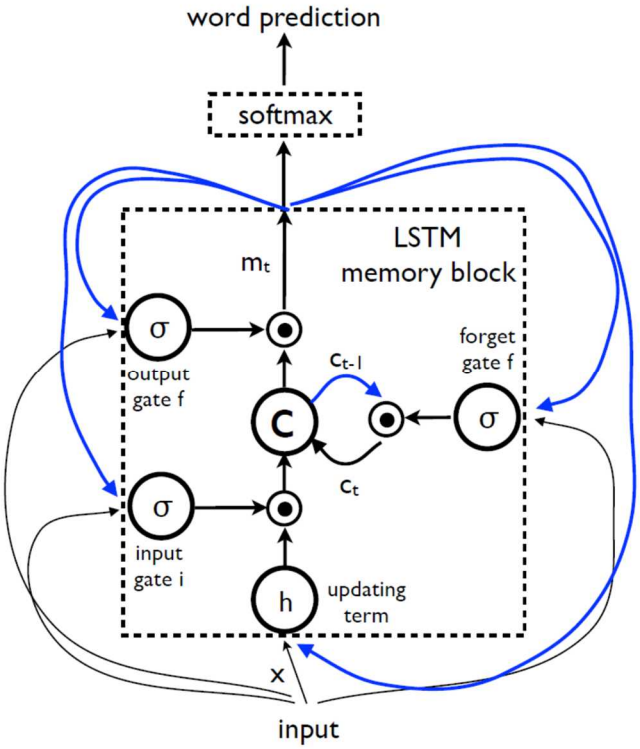


図２

上記図２は、LSTM のメモリブロックを示す説明図である。メモリブロックは、３つのゲートにより制御されるセル c を含む。時間 t -1 での出力 m は、3 つのゲートを介して時間 t でメモリにフィードバックされる。当該セルの値は忘却ゲートを介してフィードバックされる。時間 t -1 で予測された単語は、時間 t でのメモリ出力 m に加えて、単語予測のための Softmax にフィードバックされる。

LSTM モデルは、 $p(S_t | I, S_0, ..., S_t\text{-}1)$で定義された先行するすべての単語だけでなく、画像も見た後に、文の各単語を予測するようにトレーニングされる。この目的のため LSTM は展開された形とする。

下記図３に示すように、すべての LSTM が同じパラメータを共有し、時間 t-1 での LSTM の出力 m_{t-1} が時刻 t で LSTM にフィードバックされるように、LSTM メモリのコピーが画像と各センテンスワードに対して生成される。

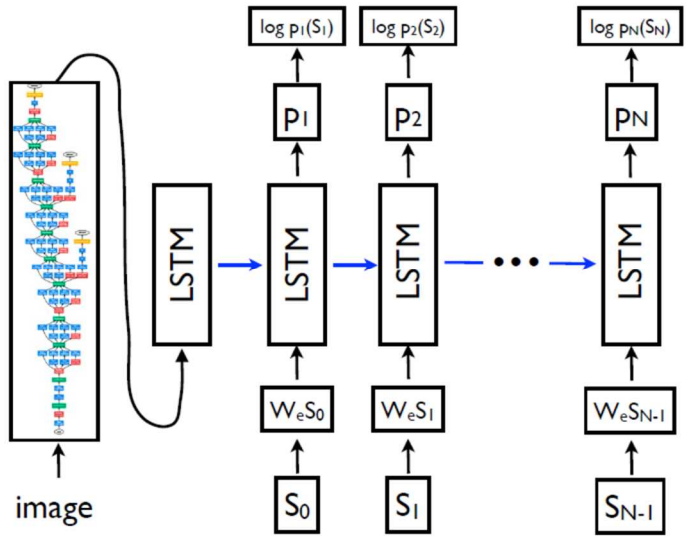


図３

図３に示すように LSTM モデルは、CNN 画像の埋め込みと単語の埋め込みを組み合わせている。LSTM メモリ間の展開された接続は青色で示しており、図 2 の繰り返し接続に対応しており、全ての LSTM は同じパラメータを共有している。

Approach	PASCAL (xfer)	Flickr 30k	Flickr 8k	SBU
Im2Text [24]				11
TreeTalk [18]				19
BabyTalk [16]	25			
Tri5Sem [11]			48	
m-RNN [21]		55	58	
MNLM [14][5]		56	51	
SOTA	25	56	58	19
NIC	**59**	**66**	**63**	**28**
Human	69	68	70	

上記テーブルは、変換されたテキストの品質を示す BLEU スコアを示す。SOTA は State of the Art の略である。NIC が本論文のニューラル画像キャプションである。人間のスコアに近い結果が示されている。下記図は NIC により生成された文章を人間が評価したものである。左側が正確に文章を生成できているグループであり、右側に行くにつれ不正確となっている。

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