

AI 特許紹介(23)
AI 特許を学ぶ！究める！
～解析ツリー生成 AI～

2020 年 12 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 12 月 30 日

登録日 2019 年 4 月 23 日

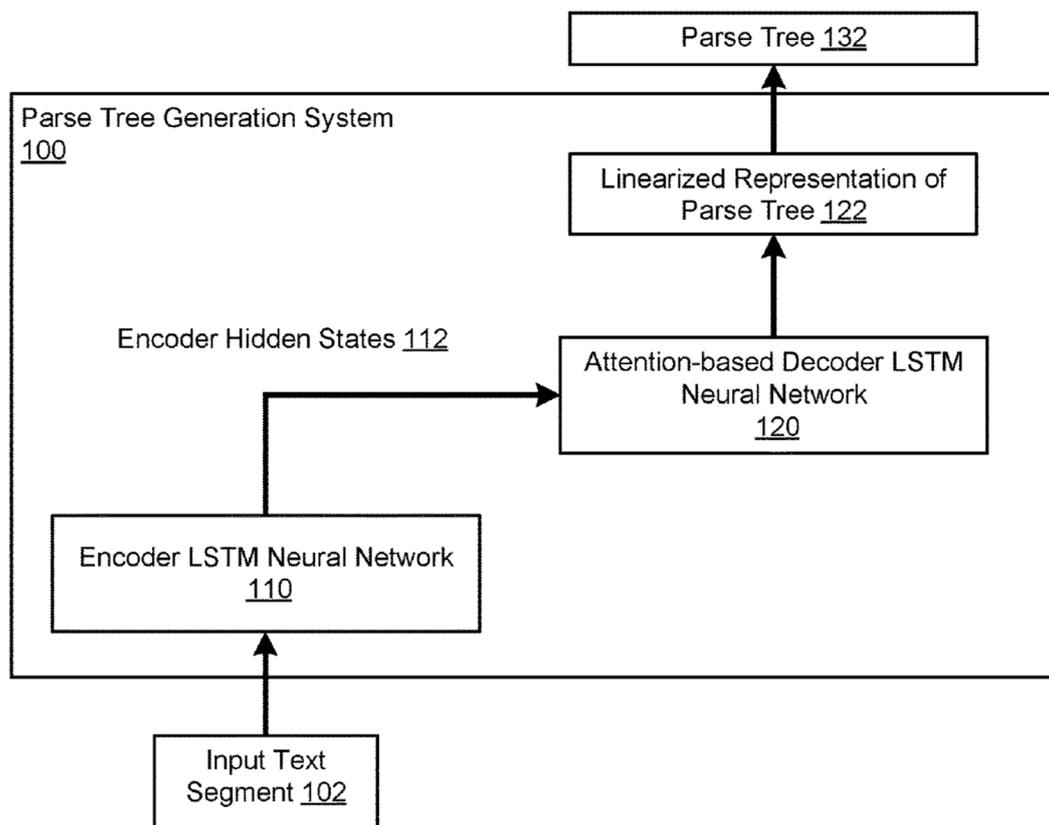
登録番号 US10268671

発明の名称 ニューラルネットワークを使用したテキストセグメントの解析ツリーの生成

671 特許は、アテンション機構を有するニューラルネットワークを用いてテキストセグメントの解析ツリーを生成する技術に関する。

2.特許内容の説明

図 1 は、構文解析ツリー生成システム 100 の例を示す。



構文解析ツリー生成システム 100 は、入力テキストセグメント 102 を受信し、入力テキストセグメント 102 ごとに、それぞれの構文解析ツリー 132 を生成する。

各入力テキストセグメント 102 は、単語のシーケンスであり、たとえば、文、文の断片、別の複数の単語のシーケンス、または、特定の自然言語の句読点である。特定のテキストセグメントの解析ツリー 131 は、文脈自由文法に従うテキストセグメントの構文構造を表すツリーである。

解析ツリー生成システム 100 によって取得される入力テキストセグメントは、様々な長さのシーケンスである。構文解析ツリー生成システム 100 は、様々な数の単語および句読点を有するテキストセグメントを受信するよう構成されている。

解析ツリー生成システム 100 は、エンコーダ長短期記憶 (LSTM) ニューラルネットワーク 110 およびアテンションベースのデコーダ LSTM ニューラルネットワーク 120 を含む。

入力テキストセグメントからの解析ツリーの生成の一部として、解析ツリー生成シ

テム 100 は、エンコーダ LSTM ニューラルネットワーク 110 を使用して入力テキストセグメントを処理し、入力テキストセグメント内の各入力に対してそれぞれのエンコーダ隠れ状態を生成する。例えば、入力テキストセグメント 102 の入力のための隠れ状態 112 のセットである。

エンコーダ LSTM ニューラルネットワーク 110 は、1 つまたは複数の 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 つまたは複数の LSTM メモリブロックを含む。

各 LSTM メモリブロックは、それぞれが入力ゲート、忘却ゲート、および出力ゲートを含む 1 つまたは複数のセルを含むことができ、これにより、セルは、現在の活性化を生成するために使用するための、または、LSTM ニューラルネットワーク 110 の他の構成要素に提供されるための隠れ状態として、セルによって生成された以前の活性化を記憶することができる。

エンコーダ LSTM ニューラルネットワーク 110 は、例えば、トレーニングを通じて、入力テキストセグメント内の各入力を処理して、パラメータのセットに従って入力テキストセグメント内の入力の隠れ状態を生成するように構成されている。

特に、エンコーダ LSTM ニューラルネットワーク 110 は、入力テキストセグメント内の各入力を順番に受信し、受信した入力を処理することにより、エンコーダ LSTM ニューラルネットワーク 110 の現在の隠れ状態を更新するよう構成されている。すなわち、現在受信した入力を処理することによって入力テキストセグメントからの以前の入力を処理することによって生成されたエンコーダ LSTM ニューラルネットワーク 110 の現在の隠れ状態を変更する。

解析ツリー生成システム 100 は、アテンションベースのデコーダ LSTM ニューラルネットワーク 120 を使用して、入力テキストセグメントのエンコーダ隠れ状態 112 を処理して、入力テキストセグメント 102 の解析ツリーの線形化表現 122 を生成する。

解析ツリーの線形化表現は、解析ツリーを集合的に表す、事前に決定された解析ツリーシンボルの語彙からのシンボルのシーケンスである。特に、構文解析ツリーシンボルの語彙には、構文解析ツリーの一部を表すシンボル、たとえば、スピーチタグの一部等、シーケンス内の他のシンボル間の階層関係を決定するシンボルなどが含まれる。シンボルはまた、未知の単語を表す未知の単語のシンボルを含む。

次に、解析ツリー生成システム 100 は、線形化表現 122 から解析ツリー132 を生成する。すなわち、線形化表現 122 内のシンボルを深さ優先走査順序で解析ツリーに追加し、シーケンス内のシンボルによって指定される階層関係に一致するようにツリー内にレベルを生成する。

解析ツリー生成システム 100 は、さらなる処理のための別のシステムへの入力として、解析ツリー132 または線形化表現 122 のいずれかを提供する。例えば、解析ツリー生成システム 100 は、入力テキストセグメント 102 を分析する自然言語処理システムへの入力として解析ツリー132 または線形化表現 122 を提供することができる。例えば、感情分析、質問応答、要約、または別の自然言語処理タスクを実行する。

アテンションベースのデコーダ LSTM ニューラルネットワーク 120 は、1 つまたは複数の LSTM 層を含み、線形化表現およびエンコーダの隠れ状態で現在のシンボルを受け取るよう構成されている。また、現在のシンボルからの可能な出力のセットのそれぞれについて、アテンションベースのデコーダ LSTM ニューラルネットワーク 120 の現在の隠れ状態およびパラメータのセットの現在の値に従って、それぞれの出力スコアを生成するよう構成されている。

アテンションベースのデコーダ LSTM ニューラルネットワークはディープ LSTM ニューラルネットワークである。特定のシンボルの出力スコアは、そのシンボルが線形化表現の次のシンボルである可能性を表す。つまり、そのシンボルは線形化表現の現在のシンボルの直後に続く。

出力スコアを生成する一部として、アテンションベースのデコーダ LSTM ニューラルネットワーク 120 は、ネットワークの隠れ状態を更新して、更新された隠れ状態を生成する。一般に、出力順序の各位置について、アテンションベースの LSTM ニューラルネットワーク 120 は、現在のシンボルおよびエンコーダの隠れ状態を処理して、入力エンコーダ隠れ状態に対してアテンションベクトルを生成し、位置の出力スコアのセットを生成すべく、アテンションベクトルを処理するように構成される。

次に、構文解析ツリー生成システム 100 は、出力スコアに従って、すなわち、出力スコアが最も高いシンボルを選択することによって、出力順序における位置のシンボルを選択する。

3.クレーム

671 特許のクレーム 1 は以下の通りである。

1. 1つまたは複数のコンピュータによって実行される方法であつて、以下を含む方法；
 入力順序に従って配置された複数の入力を含む入力テキストセグメントを取得し、
 入力テキストセグメントの各入力に対してそれぞれのエンコーダ隠れ状態を生成するために、エンコーダ長短期記憶 (LSTM) ニューラルネットワークを使用して入力テキストセグメントの入力を処理し、

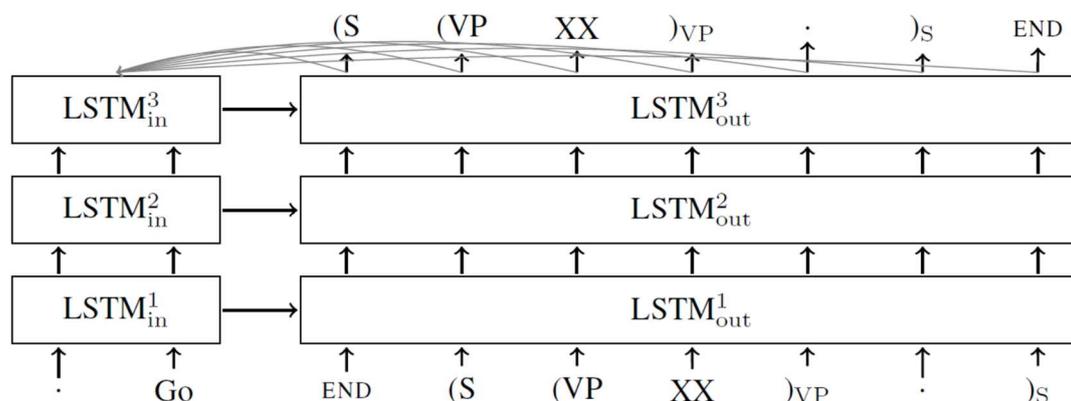
入力テキストセグメントの解析ツリーの線形化表現を生成するために、アテンションベースのデコーダ LSTM ニューラルネットワークを使用して、入力テキストセグメントの入力のそれぞれのエンコーダ隠れ状態を処理し、

ここで、入力テキストセグメントの解析ツリーの線形化表現は、出力順序に従って配置された解析ツリーシンボルの所定の語彙からのシンボルのシーケンスであり、各入力は、シンボルのシーケンスからの1つまたは複数のシンボルに対応し、

ここで、エンコーダ LSTM ニューラルネットワークおよびアテンションベースのデコーダ LSTM ニューラルネットワークは、トレーニングテキストセグメントのセット、および、各トレーニングテキストセグメントについて、トレーニングテキストセグメントの既知の解析ツリーのそれぞれの既知の線形化表現を含むトレーニングデータに対する時間トレーニング技術による逆伝播を使用してトレーニングされており、それぞれの既知の線形化表現は、既知の解析ツリーを表すシンボルの所定の語彙からのシンボルのトレーニングシーケンスである。

4. アテンション機構に関する論文

本特許に関連するアテンション機構に関する論文¹が Oriol Vinyals 氏らにより発表されている。



¹ Oriol Vinyals, Lukasz Kaiser, Terry Koo, Slav Petrov, Ilya Sutskever, Geoffrey Hinton "Grammar as a Foreign Language" arXiv:1412.7449v3 [cs.CL] 9 Jun 2015

上記図は、「Go.」という文での LSTM + A モデルの実行の概略図である。

Sequence-to-Sequence モデルに、アテンション機構を追加している。各出力シンボル B_t を生成するために、エンコーダの LSTM 状態に対してアテンション機構を使用するアテンションモデルを採用している。具体的には図に示すように、2つの別々の LSTM を使用する（1つは入力ワード A_i のシーケンスをエンコードし、もう1つは出力シンボル B_i を生成またはデコードする）。

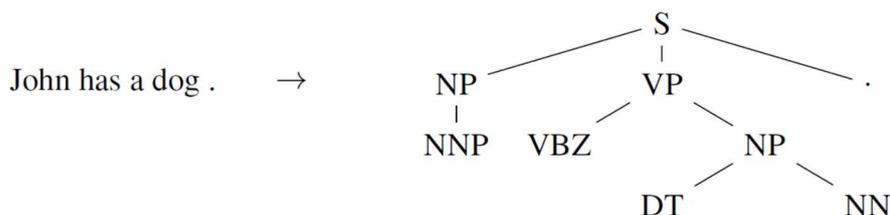
エンコーダの隠れ状態は (h_1, \dots, h_{T_A}) で示され、デコーダの隠れ状態は (d_1, \dots, d_{T_B}) : $= (h_{T_A+1}, \dots, h_{T_A+T_B})$ で示される。入力単語 $(1, \dots, T_A)$ の各出力時間 t でアテンションベクトルを計算すべく、以下のとおり定義する。

$$u_i^t = v^T \tanh(W_1' h_i + W_2' d_t)$$

$$a_i^t = \text{softmax}(u_i^t)$$

$$d_t' = \sum_{i=1}^{T_A} a_i^t h_i$$

ベクトル v と行列 W_1', W_2' は、モデルの学習可能なパラメータである。ベクトル u^t の長さは T_A であり、その i 番目の項目には i 番目の隠れエンコーダ状態 h_i にどれだけ注意を払うべきかというスコアが含まれている。これらのスコアは **softmax** によって正規化され、エンコーダの隠れ状態に対しアテンションマスクを生成する。最後に、 d_t' を d_t と連結する。これは、予測を行うための新しい隠れ状態になり、リカレントモデルの次のタイムステップに送られる。



John has a dog . → $(S (NP NNP)_{NP} (VP VBZ (NP DT NN)_{NP})_{VP} .)_{S}$

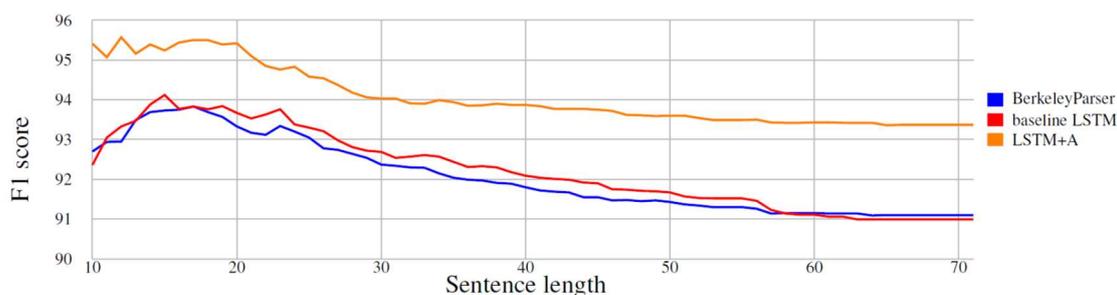
上記図は、解析ツリー及びその線形化表現を示す。

上記のモデルを解析に適用するには、解析ツリーをシーケンスに変換する可逆的な方法を設計する必要がある（線形化）。深さ優先探索の順序に従い、最初にネットワークは左から右へのスweepで文を消費し、メモリ内にベクトルを作成する。次いで、これ

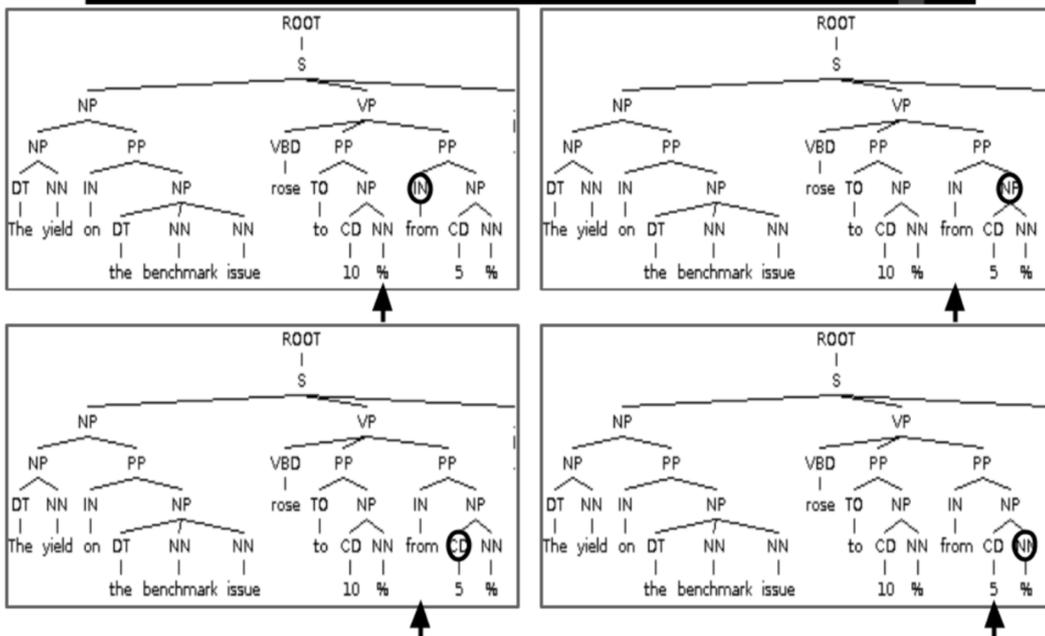
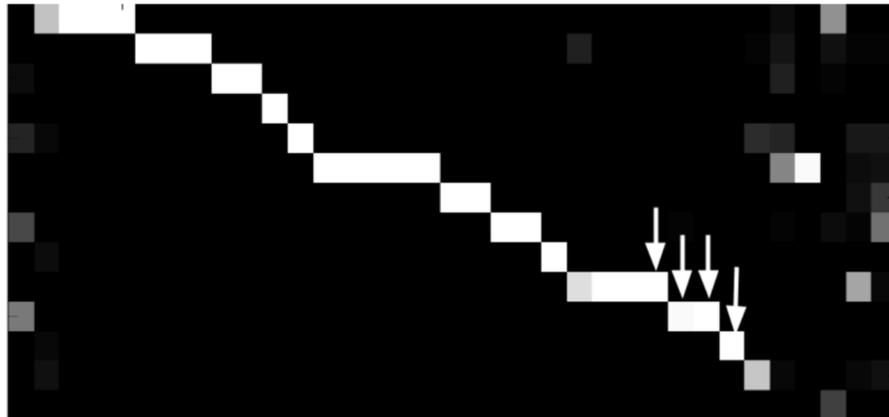
らのベクトルの情報を使用して、線形化された解析ツリーを出力する。3つの LSTM レイヤーを使用し、入力文を逆にして、品詞タグを正規化する。「Go」という文での LSTM +A モデルの実行例が上記図に示されている（上部の灰色のエッジはアテンションを示している）。

Parser	Training Set	WSJ 22	WSJ 23
baseline LSTM+D	WSJ only	< 70	< 70
LSTM+A+D	WSJ only	88.7	88.3
LSTM+A+D ensemble	WSJ only	90.7	90.5
baseline LSTM	BerkeleyParser corpus	91.0	90.5
LSTM+A	high-confidence corpus	93.3	92.5
LSTM+A ensemble	high-confidence corpus	93.5	92.8

上記テーブルは F1 スコアを示している。通常の LSTM と比較して LSTM+A (LSTM にアテンション機構を追加) 及び LSTM+A+D (D:ドロップアウト) が高いスコアを記録している。



また文が長くなったとしても、F1 スコアは高いスコアを維持している。



上側に示されているのはアテンションマトリックスで、各列は入力に対するアテンションベクトルを示す。文章は”The yield on the benchmark issue rose to 10% from 5%”である。下側には、アテンションマスクが右に移動する、4つの連続するタイムステップの出力が表示されている。図に示す通り、ターミナルノードが消費されるたびに、アテンションポインタが右に移動する。なお、黒矢印はモデルがアテンンドする場所であり、黒丸がツリー内でデコードされている現在の出力である。

各列が入力に対するアテンションベクトルであるアテンションマトリックスから、モデルが解析ツリーを生成する際に 1 つの単語に非常に鋭く焦点を合わせていることが理解できる。また、フォーカスが最初の単語から最後の単語に単調に移動し、単語が消費されると決定論的に右に移動することも明らかである。

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「世界のソフトウェア特許(共著)」、「FinTech 特許入門」、「[AI/IoT 特許入門 2.0](#)」がある。