AI 特許紹介(35)
AI 特許を学ぶ！究める！
～Universal Transformer～

2021 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　2019 年 5 月 20 日
登録日　2020 年 8 月 11 日
登録番号 US10740433
発明の名称　Universal Transformers

433 特許は、セルフアテンション機構を有する Transformer に、全てタイムステップにわたってリカレント遷移関数を適用した Universal Transformer に関する。

2.特許内容の説明
2017 年に Vaswani らにより発表された Transformer 等のセルフアテンション型フィードフォワードシーケンスモデルは、機械翻訳、画像生成、構成要素解析などのシーケンスモデリングタスクで印象的な結果を達成することが示されており、多くのシーケンスのモデリング問題のためのデファクトスタンダードアーキテクチャであるリカレントニューラルネットワークの魅力的な代替手段となっている。

しかしながら、Transformer は、リカレントモデルが簡単に処理できる一部のタスクを一般化することができない。これには、文字列または数式の長さがトレーニング時に観察された長さを超える場合の、文字列のコピーまたは単純な論理的推論が含まれる。

一般的なシーケンスに合わせたリカレントの代わりに、Universal Transformer は、シーケンスのさまざまな部分からの情報を組み合わせるためにセルフアテンションを採用しながら、深さ方向に反復する。

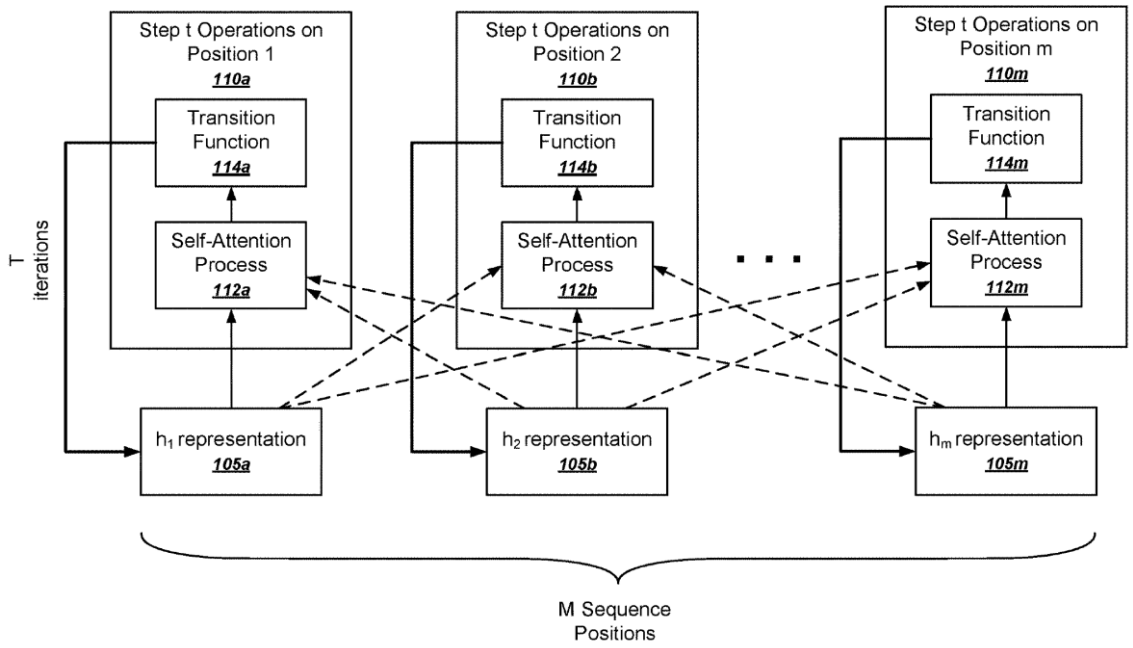


図１　シーケンスに対する例示的な Universal Transformer の動作を示す図

図１は、シーケンスに対する Universal Transformer の動作を示す図である。
図１に示される計算構造は、下記図４に示すエンコーダコンピュータシステムまたはデコーダコンピュータシステム上に実装される。

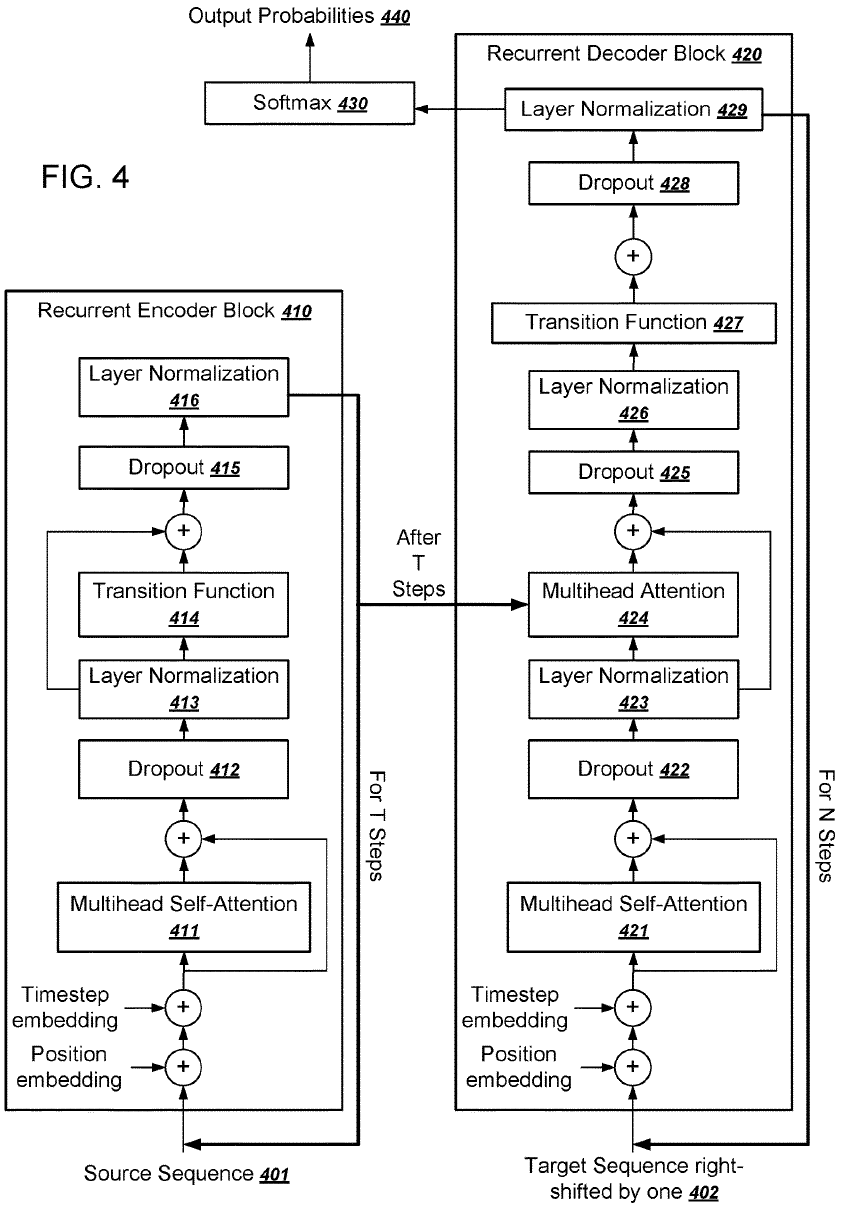


FIG. 4

図 4　Universal Transformer の詳細スキーム図

Universal Transformer の計算構造は、並列処理システムとして実装され、並列処理システムの各計算リソースは、シーケンス内の 1 つ以上の位置の操作を実行する。

一般に、エンコーダまたはデコーダのいずれかを実装するために、システムは、T 回の反復で、M 系列の位置に対して、並列に同じ一連のエンコードまたはデコード操作を実行する。

各位置での各タイムステップの操作には、少なくともセルフアテンションプロセスと遷移機能(transition function)を含めることができる。例えば、ステップ t において、システムは、セルフアテンションプロセス 112a および遷移関数 114a を使用して、シーケンスの最初の要素、h_1 表現 105a を処理する。

次に、システムは、h_1 表現 105a を更新し、T 反復に対して同じステップを繰り返す。同様に、システムは、セルフアテンションプロセス 112b とそれに続く遷移関数 114b および h_2 表現 105b の更新を使用して、T 反復のために、シーケンスの第 2 の要素、h_2 表現 105b を処理する。

同様に、システムは、セルフアテンションプロセス 112m とそれに続く遷移関数 114m および h_m 表現 105m の更新を使用して、T 反復のために、シーケンスの最後の要素である h_m 表現 105m を処理する。

図 1 の点線で示されるように、各シーケンス位置のセルフアテンションプロセス 112a～m は、他のシーケンス位置の現在の表現を入力として使用する。言い換えると、各ステップで、システムは複数の他の位置に対してこれまでに生成された表現の特定の位置の出力を調整することができる。

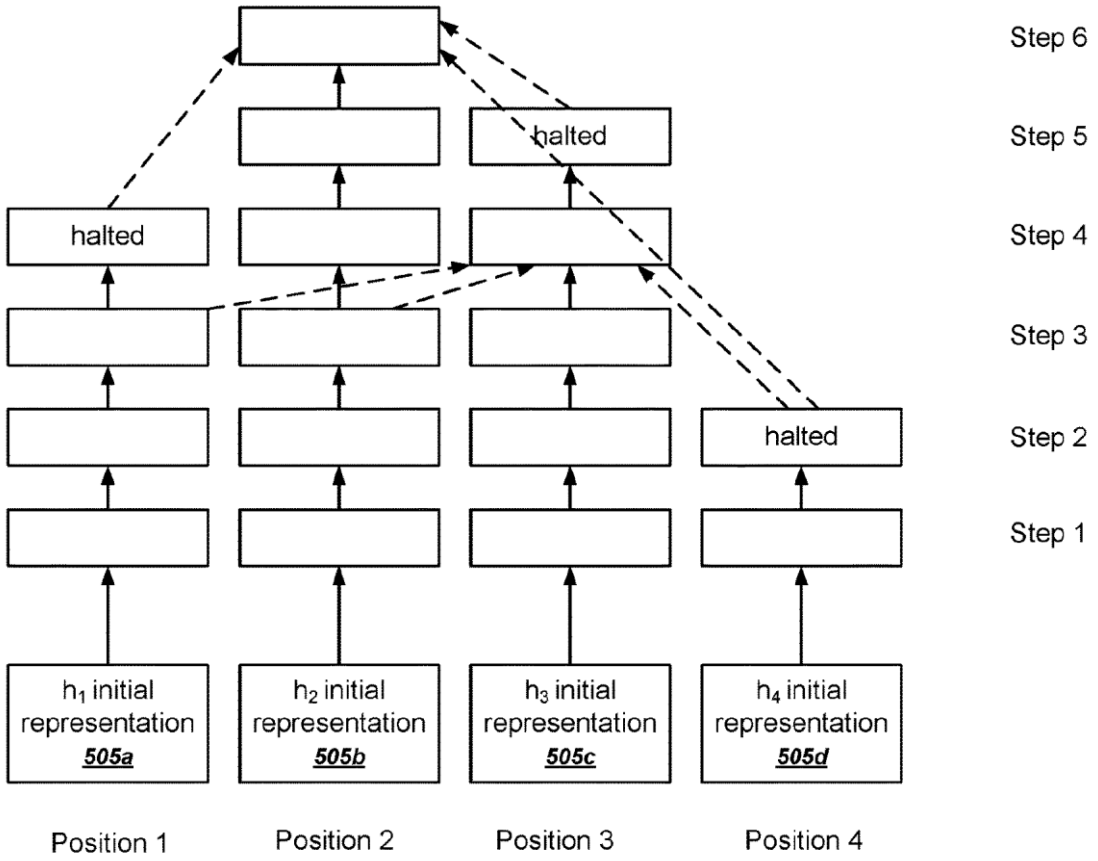


図５　シーケンス内の要素ごとのダイナミック選択

図５は、シーケンス内の要素ごとのいくつかの処理ステップのダイナミック選択を示している。シーケンスは４つの要素を有し、それぞれがそれぞれの初期表現 505a、505b、505c、および 505d を有する。

ステップ 1 で、システムは、各位置のそれぞれのエンコーダまたはデコーダーブロックを使用して、4 つの修正された表現を並行して生成する。
ステップ 2 で、位置 4 のブロックが停止し、システムは他の位置の 3 つの修正された表現を生成する。
ステップ 3 で、システムはまだ停止していない 3 つの位置に対して 3 つの修正された表現を生成する。
ステップ 4 で、位置 1 のブロックが停止する。システムは、位置 2 と位置 3 で停止していない他の位置に対して 2 つの修正された表現を生成する。

図 5 の破線は、修正された表現を生成するために使用されるセルフアテンションプロセスへの入力を表す。ステップ４で、システムは、セルフアテンションプロセスへの入

力として他の位置の表現を使用することによって、位置 3 の修正された表現を生成する。

特に、図 5 に示すようにシステムは、ステップ 2 の位置 4 に停止表現を使用し、位置 1 と 2 にはステップ 3 の表現を使用する。つまり、システムは異なる時間に異なるステップで生成された表現を使用して、セルフアテンションプロセスを実行する。

ステップ 5 で、位置 3 のブロックが停止し、システムは位置 2 に対してのみ修正された表現を生成する。
ステップ 6 で、システムは 3 つの異なるステップで生成された表現（位置 3 のステップ 5 からの表現、位置 1 のステップ 4 からの表現、および位置 4 のステップ 2 からの表現）を使用して、位置 2 の修正された表現を生成する。プロセスの最終出力は、各位置で停止した最終表現のコレクションである。

3.クレーム

433 特許のクレーム 1 は以下の通りである。

1. 1 台以上のコンピュータによって実装されるシステムにおいて、

それぞれがそれぞれの初期入力表現を有する要素の入力シーケンスを受信し、エンコーディング処理の複数の時間ステップのそれぞれについて、シーケンスのすべての要素に同じ一連のエンコーディング操作を並行して繰り返し適用することによって入力表現を修正するように構成されたエンコーダを備え、前記エンコーディング処理は、最大で所定の最大時間ステップ数について、エンコーディング処理の複数の時間ステップにおける各時間ステップで要素の表現を修正することを含み、

シンボルのターゲットシーケンス y=(y1, . . . , yn)を自己回帰的にデコードするように構成されており、デコードプロセスの複数のタイムステップのすべてのタイムステップで、デコードプロセスの前のシンボルとシーケンスのエンコーダの最終出力を条件付けるデコーダと

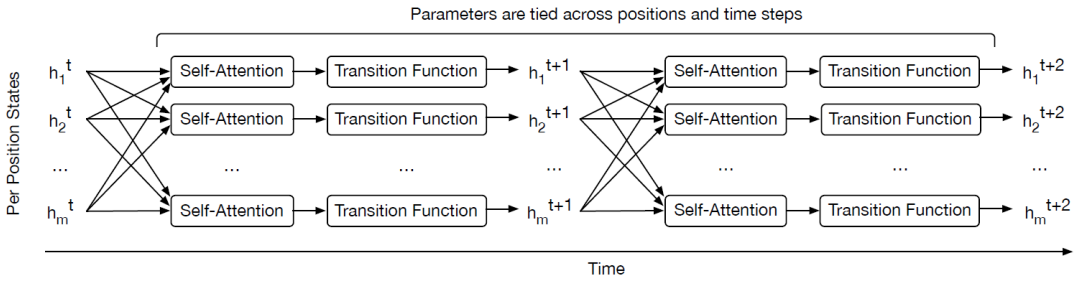
を備える。

4. 本特許に関する論文

本特許に関する論文“UNIVERSAL TRANSFORMERS”が、Dehghani らにより公表されている[1]。

[1] Mostafa Dehghani, Stephan Gouws, Oriol Vinyals, Jakob Uszkoreit, Łukasz Kaiser “UNIVERSAL TRANSFORMERS” arXiv:1807.03819v3 [cs.CL] 5Mar 2019

Universal Transformer は、Transformer のようなフィードフォワードシーケンスモデルの並列化可能性と RNN のリカレント誘導バイアスとを組み合わせる。下記図は Universal Transformer のネットワーク構造である。



Universal Transformer は、セルフアテンションを使用して異なる位置からの情報を組み合わせ（式(1)-(3)）、すべてのタイムステップ 1 <t <T にわたってリカレント遷移関数を適用することにより（式(4)(5)）、シーケンスの各位置の一連のベクトル表現を並列に繰り返し改良する。

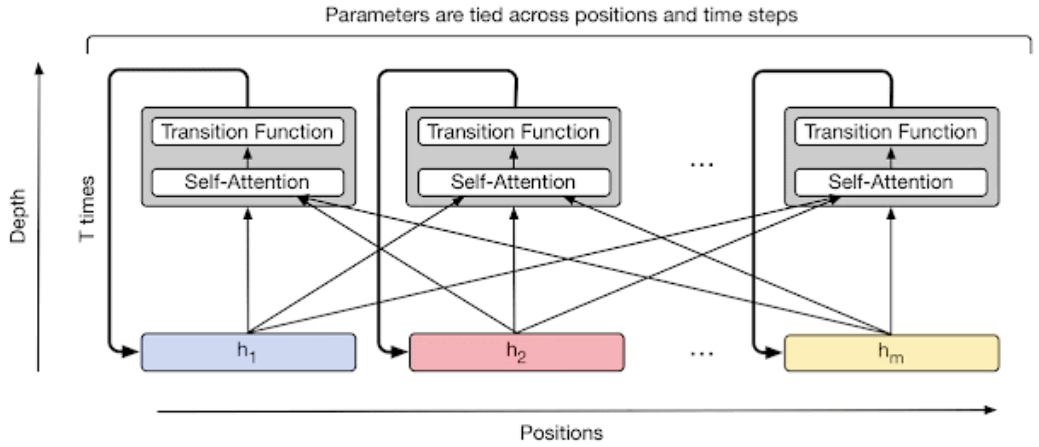
$$\text{ATTENTION}(Q,K,V) = \text{SOFTMAX}\left(\frac{QK^T}{\sqrt{d}}\right)V, \tag{1}$$

$$\text{MULTIHEADSELFATTENTION}(H^t) = \text{CONCAT}(\text{head}_1,...,\text{head}_k)W^O \tag{2}$$

$$\text{where head}_i = \text{ATTENTION}(H^tW_i^Q, H^tW_i^K, H^tW_i^V) \tag{3}$$

$$H^t = \text{LAYERNORM}(A^t + \text{TRANSITION}(A^t)) \tag{4}$$

$$\text{where } A^t = \text{LAYERNORM}((H^{t-1}+P^t) + \text{MULTIHEADSELFATTENTION}(H^{t-1}+P^t)), \tag{5}$$



具体的な動作は Google AI Blog の紹介ページの上記図が理解しやすい。
https://ai.googleblog.com/2018/08/moving-beyond-translation-with.html

質問応答タスクの目標は、複数の裏付けとなる可能性のある事実をエンコードする多数の英語の文を指定して質問に回答することである。各ストーリーで提示されている言語的事実について特定のタイプの推論を要求することにより、さまざまな形式の言語理解を測定する。

標準の Transformer は、このタイプのタスクでは良い結果を達成しない。入力をエンコードするために、システムは最初に、学習した乗法位置マスクを各単語の埋め込みに適用し、すべての埋め込みを合計することによって、ストーリー内の各事実をエンコードする。モデルは、各タスクで個別にトレーニングする（「シングルトレーニング」）か、すべてのタスクで、共同でトレーニングする（「ジョイントトレーニング」）。

システムは同じ方法で質問を埋め込み、Universal Transformer にこれらの事実と質問の埋め込みを供給する。さまざまな初期化を行い、検証セットのパフォーマンスに基づいて最適なモデルを使用して 10 回以上実行すると、Universal Transformer とダイナミック停止を備えた Universal Transformer の両方で、平均エラーと失敗したタスクの数の観点で、すべてのタスクで最先端の結果が得られた。 表 1 は、結果をまとめたものである。

Model	10K examples		1K examples	
	train single	train joint	train single	train joint
Previous best results:				
QRNet (Seo et al., 2016)	0.3 (0/20)	-	-	-
Sparse DNC (Rae et al., 2016)	-	2.9 (1/20)	-	-
GA+MAGE Dhingra et al. (2017)	-	-	8.7 (5/20)	-
MemN2N Sukhbaatar et al. (2015)	-	-	-	12.4 (11/20)
Our Results:				
Transformer (Vaswani et al., 2017)	15.2 (10/20)	22.1 (12/20)	21.8 (5/20)	26.8 (14/20)
Universal Transformer (this work)	0.23 (0/20)	0.47 (0/20)	5.31 (5/20)	8.50 (8/20)
UT w/ dynamic halting (this work)	**0.21 (0/20)**	**0.29 (0/20)**	**4.55 (3/20)**	**7.78 (5/20)**

下記図は bAbI タスクのアテンションをビジュアライゼーションしたものである。アテンション重みのビジュアライゼーションは、ストーリーと質問のすべての事実に関するさまざまなヘッドに基づいて、様々な時間ステップにわたって行われる。左側の異なるカラーバーは、異なるヘッド（合計 4 つのヘッド）に基づくアテンション重みを示す。

Story:

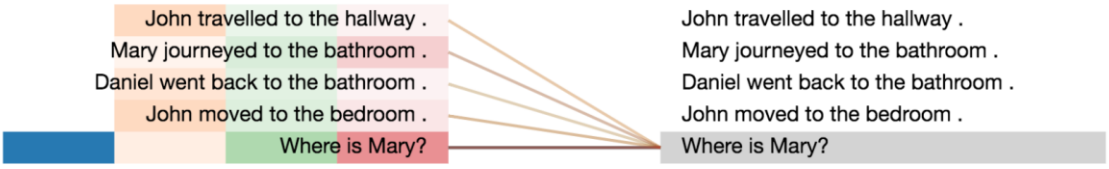
```
John travelled to the hallway.
Mary journeyed to the bathroom.
Daniel went back to the bathroom.
John moved to the bedroom
```

Question:

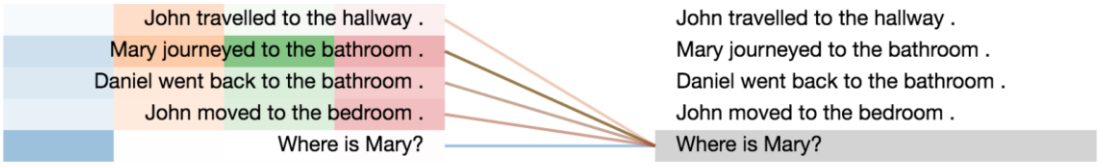
```
Where is Mary?
```

Model's output:

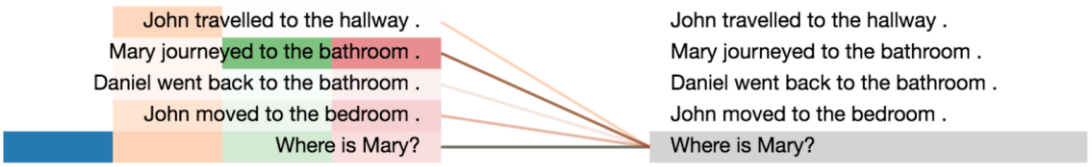
```
bathroom
```



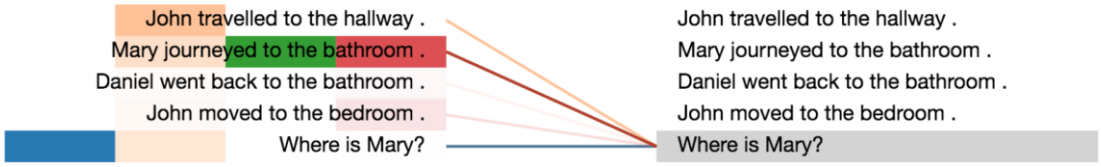
(a) Step 1



(b) Step 2



(c) Step 3



(d) Step 4

エンコードされた質問”Where is Mary?”に対して bathroom がモデルから出力される。

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」、「ブロックチェーン 3.0(共著)」がある。