AI 特許紹介(37)
AI 特許を学ぶ！究める！
～スロットアテンション特許～

2022 年 2 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
特許出願人　Google
出願日　2020 年 7 月 13 日
公開日　2021 年 12 月 9 日
公開番号 US20210383199
発明の名称　スロットアテンションを用いたオブジェクト中心学習

199 特許は、トランスフォーマーに利用されるアテンション機構を活用したスロットアテンション技術に関し、具体的には、畳み込みニューラルネットワークの出力等の知覚表現に結合し、スロットと呼ばれるタスク依存の抽象的表現のセットを生成するアーキテクチャコンポーネントであるスロットアテンション技術に関する。

2.特許内容の説明
図３は、スロットアテンションモジュール 300 のブロック図を示す。

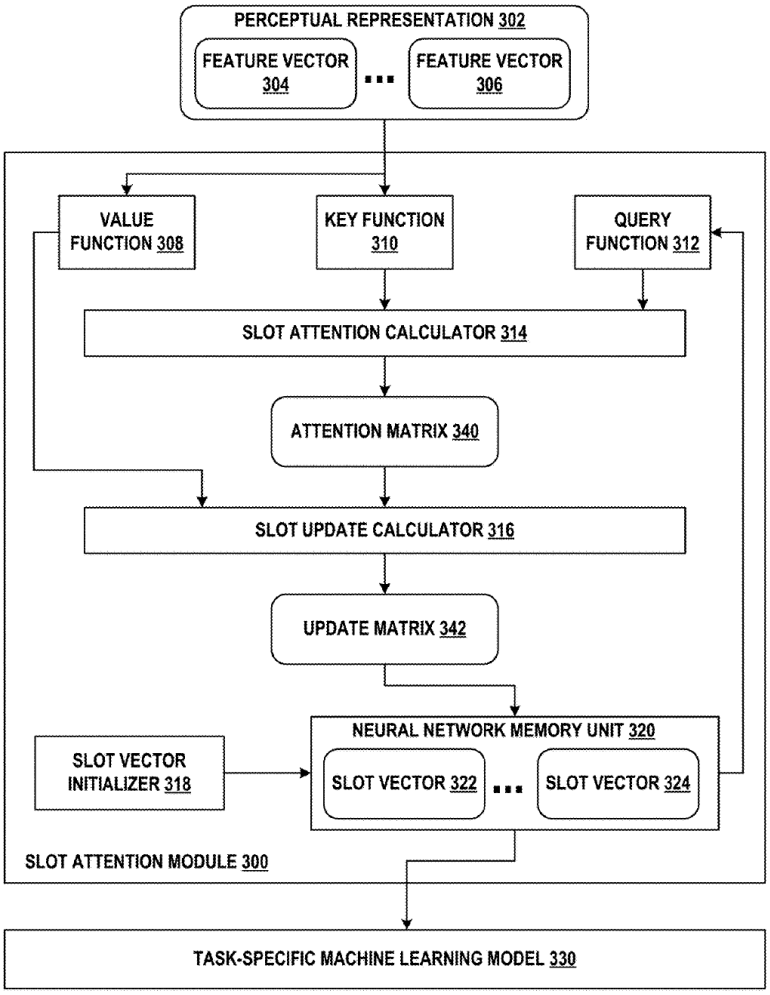


図 3

スロットアテンションモジュール 300 は、値関数 308、キー関数 310、クエリ関数 312、スロットアテンション計算機 314、スロット更新計算機 316、スロットベクトル初期化器 318、およびニューラルネットワークメモリユニット 310 を含む 。

スロットアテンションモジュール 300 は、特徴ベクトル 304～306 を含む入力知覚表現 302 を受け取る。スロットアテンションモジュール 300 は、知覚表現 302 に基づいてスロットベクトル 322～324 を生成する。

知覚表現 302 は、例えば、二次元画像データ、深度画像データ、ポイントクラウドデータ、音声データ、時系列データ、テキストデータを含む。

特徴ベクトル 304～306 の各特徴ベクトルは、複数の値を含み、各値は、特徴ベクトルの特定の次元に対応する。知覚表現 302 が画像である場合、特徴ベクトル 304～306

のそれぞれは、画像内のピクセルに関連付けられ、ピクセルの様々な特徴を表すことができる。知覚表現 302 を処理するために使用される機械学習モデルには畳み込みニューラルネットワークが含まれる。特徴ベクトル 304～306 は、知覚表現 302 の畳み込み特徴のマップを表すことができ、様々な畳み込みフィルタの出力を含む。

特徴ベクトル 304～306 のそれぞれの各特徴ベクトルは、各特徴ベクトルによって表される知覚表現 302 の一部を示す位置埋め込みを含む。特徴ベクトル 304～306 は、例えば、知覚表現 302 から抽出された畳み込み特徴に位置埋め込みを追加することによって決定することができる。それぞれの特徴ベクトルがスロットアテンションモジュール 300 に提供される順序ではなく、それぞれの特徴ベクトルの一部として特徴ベクトル 304～306 のそれぞれの各特徴ベクトルに関連する位置を符号化することにより、特徴ベクトル 304～306 を複数の異なる順序でスロットアテンションモジュール 300 に提供する。

したがって、特徴ベクトル 304～306 の一部として位置埋め込みを含めることにより、スロットアテンションモジュール 300 によって生成されたスロットベクトル 322～324 が、特徴ベクトル 304～306 に関して順列不変であることが可能になる。

画像の場合、例えば、位置埋め込みは、W×H×４テンソルを構築することによって生成することができ、ここで、WおよびHは、知覚表現 302 の畳み込み特徴のマップの幅および高さをそれぞれ表す。W×Hマップに沿ったそれぞれのピクセルに関連付けられた４つの値のそれぞれは、画像の対応する方向(上、下、右、左)に沿った画像の境界、境界、および／またはエッジに対するそれぞれのピクセルの位置を表す。

４つの値のそれぞれが 0 から 1 までの範囲に正規化される。W×H×4 テンソルは、学習可能な線形マップを使用して、畳み込み特徴と同じ次元（つまり、特徴ベクトル 304～306 と同じ次元）に投影できる。次に、投影されたW×H×４テンソルを畳み込み特徴に追加して、特徴ベクトル 304～306 を生成し、それにより、特徴ベクトル 304～306 に位置情報を埋め込むことができる。投影されたW×H×４テンソルと畳み込み特徴の合計は、１つ以上の機械学習モデル（例えば、１つ以上の多層パーセプトロン）によって処理されて、特徴ベクトル 304～306 を生成する。

特徴ベクトル 304～306 は、キー関数 310 への入力として提供される。特徴ベクトル 304～306 は、それぞれがＩ次元を有するＮ個のベクトルを含む。したがって、特徴ベクトル 304～306 は、N行（それぞれ特定の特徴ベクトルに対応する）およびＩ列を有する入力行列Ｘによって表すことができる。

キー関数 310 は、I 行およびD列を有するキー重み行列W_{KEY}によって表される線形変換を含む。キー関数 310（例えば、キー重み行列W_{KEY}）は、スロットアテンションモジュール 300 のトレーニング中に学習される。入力行列Xは、キー関数 310 によって変換されて、スロットアテンション計算機 314 への入力として提供されるキー入力行列X_{KEY}（例えば、X_{KEY} ＝ X_{WKEY}）を生成する。キー入力行列 X_{KEY} は、N 行と D 列を含む。

特徴ベクトル 304〜306 はまた、バリュー関数 308 への入力として提供される。バリュー関数 308 は、 I 行およびD列を有するバリュー重み行列W_{VERE}によって表される線形変換を含む。バリュー関数 308（例えば、値重み行列W_{VERE}）は、スロットアテンションモジュール 300 のトレーニング中に学習される。入力行列Xは、バリュー関数 308 によって変換されて、スロット更新計算機 316 への入力として提供されるバリュー入力行列X_{VERTE}（例えば、X_{VERE} ＝ X_{WVALUE}）を生成する。

キー重み行列 W_{KEY} およびバリュー重み行列 W_{VALUE} の次元は、特徴ベクトル 304〜306 の数 N に依存しないため、トレーニング中およびスロットアテンションモジュール 300 のテスト/使用中に異なる値の N を使用できる。

スロットベクトル初期化器 318 は、ニューラルネットワークメモリユニット 320 によって記憶されたスロットベクトル 322〜324 のそれぞれを初期化するように構成される。スロットベクトル初期化器 318 は、例えば、正規（すなわち、ガウス）分布から選択されたランダム値でスロットベクトル 322〜324 のそれぞれを初期化するように構成される。

スロットベクトル 322〜324 は、それぞれがＳ次元を有するＫ個のベクトルを含む。スロットベクトル 322〜324 は、K行（それぞれが特定のスロットベクトルに対応する）およびＳ列を有する出力行列Ｙによって表される。

クエリ関数 312 は、Ｓ行およびD列を有するクエリ重み行列W_{QUERY}によって表される線形変換を含む。クエリ関数 312（例えば、クエリ重み行列W_{QUERY}）は、スロットアテンションモジュール 300 のトレーニング中に学習される。

出力行列Ｙは、クエリ関数 312 によって変換されて、クエリ入力行列Y_{QUERY}（例えば、Y_{QUERY} ＝ Y_{WQURER}）を生成し、スロットアテンション計算機 314 への入力として提供される。クエリ出力行列 Y_{QUERY} は、K 行と D 列を含む。次元Ｄは、バリュー

関数 308、キー関数 310、およびクエリ関数 312 によって共有される。

さらに、クエリ重み行列 W_{QUERY} の次元はスロットベクトル 322-324 の数 K に依存しないため、トレーニング中およびスロットアテンションモジュール 300 のテスト/使用中に異なる値の K を使用できる。

スロットアテンション計算機 314 は、キー関数 310 によって生成されたキー入力行列X_{KEY}およびクエリ関数 312 によって生成されたクエリ入力行列Y_{QUERY}に基づいてアテンション行列 340 を決定する。具体的には、スロットアテンション計算機 314 は、キー入力行列X_{KEY}とクエリ出力行列Y_{QUERY}の転置との間の内積を計算する。スロットアテンション計算機 314 は、ドット積をDの平方根（すなわち、W_{VERE}、W_{KEY}、および／またはW_{QUERY}行列の列の数）で除算する。

スロット注意計算機 314 は、関数Mを実装する。

$M=(1/\sqrt{D})X_{KEY}(Y_{QUERY})^T$

Mは、アテンション行列 340 の正規化されていないバージョンを表し、N行およびK列を含む。

スロットアテンション計算機 314 は、出力軸に関して（すなわち、スロットベクトル 322～324 に関して）行列Mの値を正規化することによってアテンション行列 340 を決定する。行列Mの値は、その行に沿って（すなわち、スロットベクトル 322～324 の数に対応する次元Kに沿って）正規化され、行の各値は、それぞれの行に含まれるK値に関して正規化される。

スロットアテンション計算機 314 は、それぞれの行の複数の値に関して、行列Mのそれぞれの各行の複数の値のそれぞれの各値を正規化することによって、アテンション行列 340 を決定する。

具体的には、スロットアテンション計算機 314 は、$A_{i,j}$ に従ってアテンション行列 340 を決定する。

$A_{i,j}=(e^{M\ i,j})/(\Sigma_{l=1}{}^{K}e^{M\ i,l})$

$A_{i,j}$ は、アテンション行列 340 の行 i および列 j に対応する位置での値を示す。

行列Mは、正規化の前に転置され、行列ＭＴの値は、その列に沿って（すなわち、スロットベクトル 322～324 の数に対応する次元Kに沿って）正規化される。行列ＭＴのそれぞれの各列の各値は、それぞれの列に含まれるK値に関して正規化される。

スロットアテンション計算機 314 は、$A_{i,j}$ に従ってアテンションマトリックス 340 の転置バージョンを決定する。

$A_{i,j}{}^{T}=(e^{M\ i,j})/(\Sigma_{l=1}{}^{K}e^{M\ i,j})$

$A_{i,j}{}^{T}$ は、転置されたアテンション行列 340 の行 i および列 j に対応する位置での値を示し、これは、転置されたアテンション行列ＡＴと呼ばれる。

転置されたアテンション行列 340 は、出力軸に関して（すなわち、スロットベクトル 322～324 に関して）行列Mの値を正規化することによって決定される。

スロット更新計算機 316 は、バリュー関数 308 およびアテンション行列 340 によって生成された値入力行列X_{VALUE}に基づいて更新行列 342 を決定する。スロット更新計算機 316 は、アテンション行列Ａの転置とバリュー入力行列X_{VALUE}とのドット積を決定することによって更新行列 342 を決定する。

スロット更新計算機 316 は、関数 $U_{WEIGHTED\ SUM}=A^{T}x_{VALUE}$ を実装することができ、アテンション行列Ａは、加重和計算の重みを指定するものと見なすことができ、バリュー入力行列X_{VALUE} は、加重和計算の値を指定するものと見なすことができる。更新マトリックス 342 は、Ｋ行およびＤ列を含む $U_{WEIGHTED\ SUM}$ によって表すことができる。

別の実装形態では、スロット更新計算機 316 は、アテンション重み行列 $W^{ATTENTION}$ とバリュー入力行列X_{VALUE}の転置のドット積を決定することによって更新行列 342 を決定するように構成される。

注意重み行列 $W^{ATTENTION}$ の要素/エントリは、

$W_{i,j}{}^{ATTENTION}=(A_{i,j})/(\Sigma l=1^{N}A_{i,j})$、

または、その転置の場合は

$(W_{i,j}{}^{ATTENTION})^{T}=(A_{i,j}{}^{T})/(\Sigma l=1^{N}A_{i,l}{}^{T})$

と定義できる。

したがって、スロット更新計算機 316 は、関数 $U_{WEIGHTED\ MEAN}=(W^{ATTENTION})^{T}X_{VALUE}$ を実装することができ、行列 A は、加重平均計算の重みを指定していると見なすことができ、値入力行列 X_{VALUE} は、加重平均計算の値を指定していると見なすことができる。更新マトリックス 342 は、Ｋ行およびＤ列を含む $U_{WEIGHTED\ MEAN}$ によって表すことができる。

更新行列 342 は、ニューラルネットワークメモリユニット 320 への入力として提供

され、スロットベクトル 322－324 の以前の値に基づいてスロットベクトル 322－324 を更新し、行列 342 を更新する。

ニューラルネットワークメモリユニット 320 は、ゲート付き回帰ユニット（ＧＲＵ）、長短期記憶（ＬＳＴＭ）ネットワーク、スロットベクトル 322-324 を格納または更新するように構成された他のニューラルネットワークまたは機械学習ベースのメモリユニットを含む。

ニューラルネットワークメモリユニット 320 は、各処理反復中にスロットベクトル 322～324 の一部のみを更新するのではなく、各処理反復中にスロットベクトル 322～324 のそれぞれを更新する。

スロットベクトル 322－324 の更新された値として更新行列 342 を使用するのではなく、その前の値および更新行列 342 に基づいてスロットベクトル 322－324 の値を更新するようにニューラルネットワークメモリユニット 310 を訓練することで、精度を改善し、スロットベクトル 322-324 の収束を高速化する。

スロットアテンションモジュール 300 は、反復的にスロットベクトル 322～324 を生成する。つまり、スロットベクトル 322-324 は、タスク固有の機械学習モデル 330 に入力として渡される前に、1 回以上更新される。たとえば、スロットベクトル 322-324 は、タスク固有の機械学習モデル 330 での使用にあたり「準備完了」と見なされる前に 3 回更新される。

具体的には、スロットベクトル 322－324 の初期値は、スロットベクトル初期化子 318 によってそれに割り当てられる。初期値がランダムである場合、それらは知覚表現 302 に含まれる実体を正確に表さない可能性が高い。したがって、特徴ベクトル 304～306 およびランダムに初期化されたスロットベクトル 322−324 は、スロットアテンションモジュール 300 の構成要素によって処理され、スロットベクトル 322−324 の値をリファインし、それによって更新されたスロットベクトル 322−324 を生成する。

この最初の反復またはスロットアテンションモジュール 300 を通過した後、スロットベクトル 322～324 のそれぞれは、知覚表現 302 に含まれる 1 つまたは複数の対応するエンティティに注意を向け、結合し、表現し始めることができる。

特徴ベクトル 304～306 および現在更新されているスロットベクトル 322−324 は、スロットアテンションモジュール 300 の構成要素によって再び処理されて、スロット

ベクトル 322−324 の値をさらにリファインし、それによってスロットベクトル 322−324 に対する別の更新を生成する。

この第２の反復またはスロットアテンションモジュール 300 を通過した後、スロットベクトル 322～324 のそれぞれは、強度を増加させて１つまたは複数の対応するエンティティに注意を向け、結合し続け、それにより、精度を増加させて１つまたは複数の対応するエンティティを表すことができる。

さらなる反復を実行することができ、追加の各反復は、スロットベクトル 322～324 のそれぞれが対応する１つまたは複数のエンティティを表す精度にいくらかの改善をもたらすことができる。所定の反復回数の後、スロットベクトル 322～324 は、ほぼ安定した値のセットに収束する。

図５Ａは、教師なし学習タスクへスロットアテンションモジュール 300 を適用した例を示す。

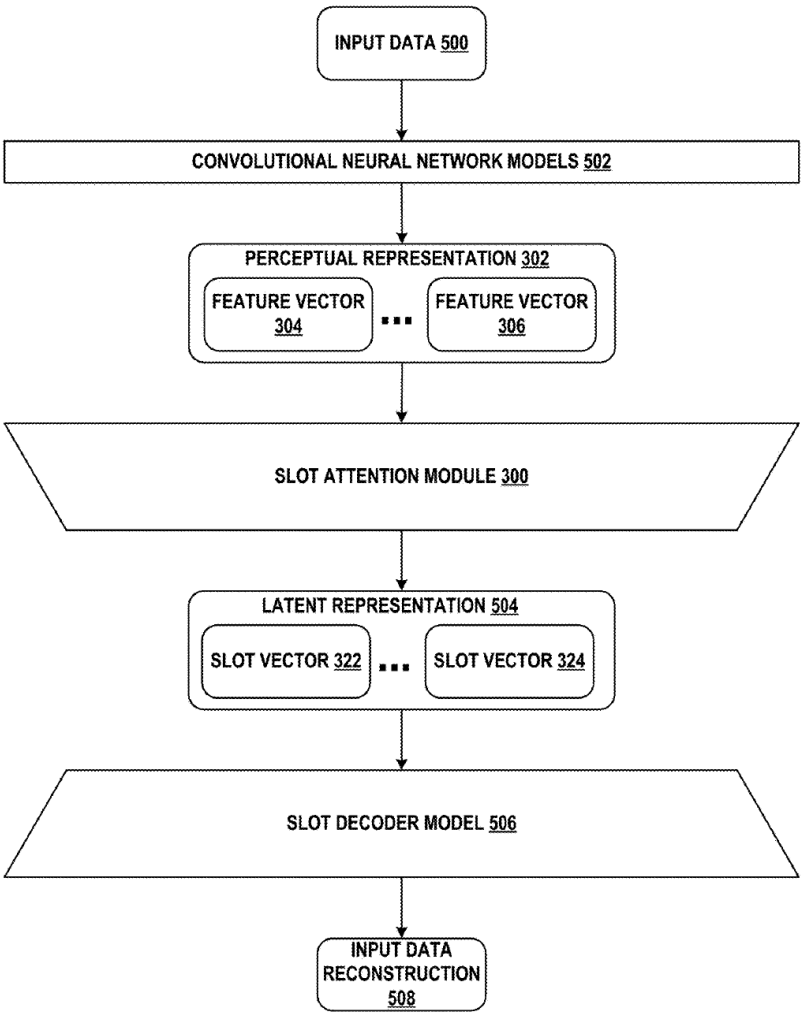


図５Ａ

畳み込みニューラルネットワークモデル 502 は、入力データ 500 に基づいて知覚表現 302 を生成する。入力データ 500 は、画像データ、時系列（例えば、波形）データ、テキストデータ、点群データ、またはボクセルデータである。

知覚表現 302 は､畳み込みニューラルネットワークモデル 502 による入力データ 500 の処理の結果を示す特徴ベクトル 304～306 を含む。

知覚表現 302 は、スロットアテンションモジュール 300 への入力として提供され、スロットアテンションモジュール 300 は、それに基づいてスロットベクトル 322～324 を生成する。

スロットデコーダモデル 506 は、スロットベクトル 322－324 を入力として受け取

り、それに基づいて、入力データ再構成 508 を生成する。スロットデコーダモデル 506 に提供されるスロットベクトル 322～324 の値は、スロットアテンションモジュール 300 による処理の 1 回または複数の反復の出力を表す。

スロットアテンションモジュール 300 およびスロットデコーダモデル 506 は、一緒に訓練され、それにより、スロットベクトル 322～324 は、入力データ 500 を再構築するためにスロットデコーダモデル 506 によって使用される入力データ 500 に存在するエンティティの埋め込みを提供する（すなわち、入力データ再構成 508 を生成する）。

スロットアテンションモジュール 300 とスロットデコーダモデル 506 とを同時に訓練することで、スロットデコーダモデル 506 がスロットベクトル 322－324 の値を「理解」することが可能となる。

図 4 は、特定の知覚表現に関して、スロットアテンションモジュール 300 による処理反復の過程で変化する複数のスロットベクトルの例を示す。

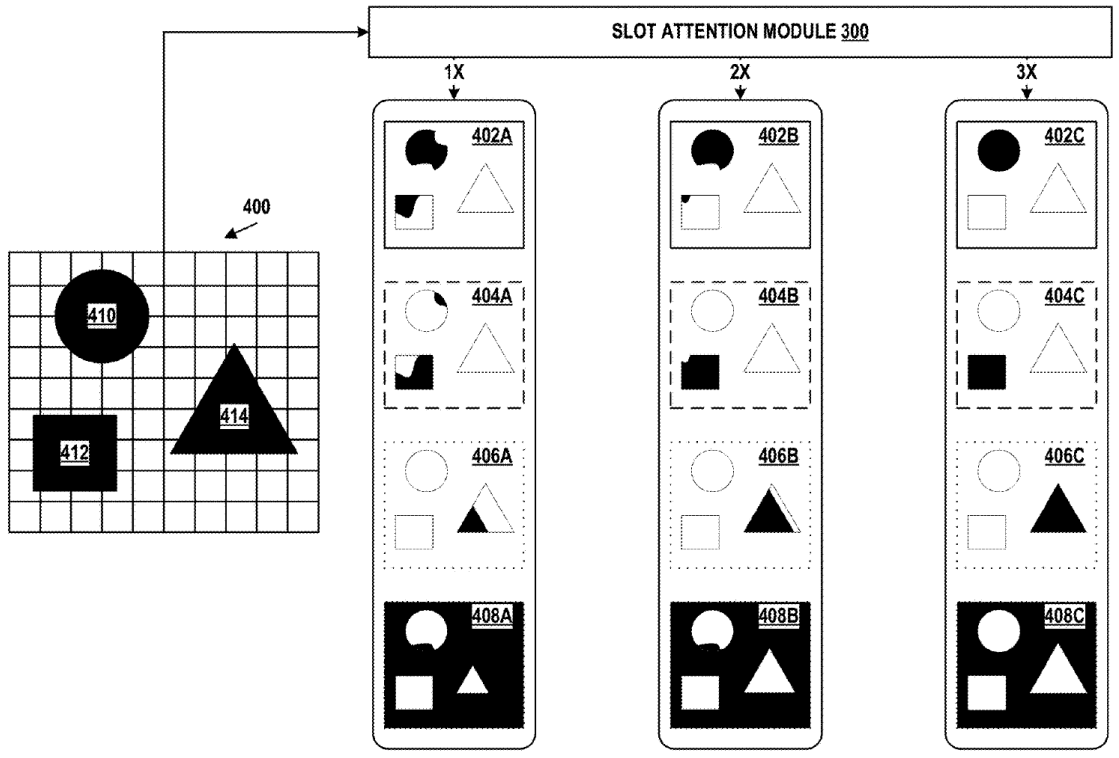


図 4

知覚表現 302 は、エンティティ 410（円形のオブジェクト）、エンティティ 412（正方形のオブジェクト）およびエンティティ 414（三角形のオブジェクト）の 3 つのエンティティを含む画像 400 によって表される。

画像 400 は、機械学習モデルによって処理され、特徴ベクトル 304～306 を生成することができ、それぞれは、画像 400 の上にオーバーレイされたグリッドの対応するグリッド要素によって表される。

グリッドの一番上の行の左端のグリッド要素は特徴ベクトル 304 を表し、グリッドの一番下の行の右端のグリッド要素は特徴ベクトル 306 を表し、それらの間のグリッド要素は他の特徴ベクトルを表す。各グリッド要素は、対応する特徴ベクトルに関連付けられた複数のベクトル値を表す。

図４では、４つのスロットベクトルを有するものとして複数のスロットベクトルを示している。ただしスロットベクトルの数は変更可能である。

スロットアテンションモジュール 300 は、画像 400 に関連する特徴ベクトルおよび４つのスロットベクトルの初期値（例えば、ランダムに初期化される）を処理して、値 402A、404A、406A、および 408A を有するスロットベクトルを生成する。

スロットベクトル値 402A、404A、406A、および 408A は、スロットアテンションモジュール 300 の第１の反復（１×）の出力を表す。スロットアテンションモジュール 300 はまた、値 402A、404A、406A、および 408A を有する特徴ベクトルおよびスロットベクトルを処理し、値 402B、404B、406B、および 408B を有するスロットベクトルを生成する。

スロットベクトル値 402B、404B、406B、および 408B は、スロットアテンションモジュール 300 の第２の反復（２×）の出力を表す。スロットアテンションモジュール 300 は、値 402B、404B、406B、および 408B を有する特徴ベクトルおよびスロットベクトルを処理し、値 402C、404C、406C、および 408C を有するスロットベクトルを生成する。

スロットベクトル値 402C、404C、406C、および 408C は、スロットアテンションモジュール 300 の第３の反復（３×）の出力を表す。スロットベクトル値 402A、404A、406A、408A、402B、404B、406B、408B、402C、404C、406C、408C の視覚化は、各反復でのアテンションマトリックス 340 に基づくアテンションマスクの視覚化、および、タスク固有の機械学習モデル 330 によって生成された再構成マスクの視覚化を表す。

値 402A、402B、および 402C に関連付けられた第 1 のスロットベクトルは、エンティティ 410 に注意を向け、結合するように構成され、それにより、エンティティ 410 の属性、プロパティ、特性を表す。

具体的には、スロットアテンションモジュール 300 の第 1 の反復後、第 1 のスロットベクトルは、スロットベクトル値 402A の視覚化における黒で塗りつぶされた領域によって示されるように、エンティティ 410 およびエンティティ 412 の態様を表すことができる。

スロットアテンションモジュール 300 の第 2 の反復の後、スロットベクトル値 402B の視覚化において、エンティティ 410 の増加した黒塗り領域およびエンティティ 412 の減少した黒塗り領域によって示されるように、エンティティ 410 のより大きな部分およびエンティティ 412 のより小さな部分を表すことができる。

スロットアテンションモジュール 300 の 3 回目の反復後、第 1 のスロットベクトルは、エンティティ 410 をほぼ排他的に表すことができ、エンティティ 410 が完全に黒で塗りつぶされ、エンティティ 412 がスロットベクトル値 402C の視覚化で完全に白で塗りつぶされていることによって示されるように、もはやエンティティ 412 を表さない。

第 1 のスロットベクトルは、スロットアテンションモジュール 300 が第 1 のスロットベクトルの値を更新または改良するときに、エンティティ 410 を表すことに収束または焦点を合わせることができる。

スロットベクトルの 1 つまたは複数のエンティティへのこのアテンションおよび収束は、スロットアテンションモジュール 300 の構成要素の数学的構造およびスロットアテンションモジュール 300 のタスク固有のトレーニングの結果である。

第 2 のスロットベクトル（値 404A、404B、および 404C に関連付けられている）は、エンティティ 412 に注意を向け、結合するように構成され、それにより、エンティティ 412 の属性、プロパティ、特性を表す。

具体的には、スロットアテンションモジュール 300 の第 1 の反復の後、第 2 のスロットベクトルは、スロットベクトル値 404A の視覚化における黒で塗りつぶされた領域によって示されるように、エンティティ 412 およびエンティティ 410 の態様を表す。

スロットアテンションモジュール 300 の第２の反復の後、第２のスロットベクトルは、エンティティ 412 のより大きな部分を表すことができ、スロットベクトル値 404B の視覚化において完全に白で塗りつぶされて示されているエンティティ 412 およびエンティティ 410 の増加した黒で塗りつぶされた領域によって示されるように、もはやエンティティ 410 を表さない。

スロットアテンションモジュール 300 の第３の反復の後、第２のスロットベクトルは、エンティティ 412 をほぼ排他的に表すことができ、スロットベクトル値 404C の視覚化において、エンティティ 412 が完全に黒で塗りつぶされ、エンティティ 410 が完全に白で塗りつぶされていることによって示されるように、もはやエンティティ 410 を表さない。

したがって、第２のスロットベクトルは、スロットアテンションモジュールが第２のスロットベクトルの値を更新または改良するときに、エンティティ 412 を表すことに収束または焦点を合わせることができる。

第３のスロットベクトル（値 406A、406B、および 406C に関連付けられる）は、エンティティ 414 に注意を向け、結合するように構成され、それにより、エンティティ 414 の属性、プロパティ、特性を表す。

具体的には、スロットアテンションモジュール 300 の第１の反復の後、第３のスロットベクトルは、スロットベクトル値 406A の視覚化における黒で塗りつぶされた領域によって示されるように、エンティティ 414 の態様を表すことができる。

スロットアテンションモジュール 300 の第２の反復の後、第３のスロットベクトルは、スロットベクトル値 404B の視覚化におけるエンティティ 414 の増加した黒で塗りつぶされた領域によって示されるように、エンティティ 414 のより大きな部分を表すことができる。

スロットアテンションモジュール 300 の第３の反復の後、第３のスロットベクトルは、スロットベクトル値 406C の視覚化において完全に黒く塗りつぶされているエンティティ 412 によって示されるように、エンティティ 414 のほぼ全体を表すことができる。

第３のスロットベクトルは、スロットアテンションモジュールが第３のスロットベクトルの値を更新または改良するときに、エンティティ 414 を表すことに収束または焦

点を合わせることができる。

第４のスロットベクトル（値 408A、408B、および 408C に関連付けられる）は、画像 400 の背景特徴に注意を向け、それに結合するように構成され、それにより、背景の属性、特性、特徴を表す。

具体的には、スロットアテンションモジュール 300 の第１の反復後、第４のスロットベクトルは、スロットベクトル値 408A の視覚化で黒く塗りつぶされた領域で示されているように、背景のほぼ全体、およびスロットベクトル値 402A 404A、または 406Aによってまだ表されていないエンティティ 410 および 414 のそれぞれの部分を表すことができる。

スロットアテンションモジュール 300 の第２の反復の後、第４のスロットベクトルは、スロットベクトル値 408B の視覚化において、背景の黒で塗りつぶされた領域およびエンティティ 410 および 414 の減少した黒で塗りつぶされた領域によって示されるように、背景のほぼ全体と、スロットベクトル値 402B、404B または 406B によってまだ表されていないエンティティ 410 および 414 のより小さな部分を表すことができる。

スロットアテンションモジュール 300 の第３の反復の後、第４のスロットベクトルは、スロットベクトル値の視覚化 408C の視覚化において、背景が完全に黒で塗りつぶされ、エンティティ 410、412、および 414 が完全に白で塗りつぶされていることによって示されるように、背景のほぼ全体をほぼ排他的に表すことができる。

したがって、第４のスロットベクトルは、スロットアテンションモジュールが第４のスロットベクトルの値を更新またはリファインするときに、画像 400 の背景を表すことに収束または焦点を合わせることができる。

3.クレーム

199 特許のクレーム１は以下の通りである。

1. コンピュータにより実装される方法において、

複数の特徴ベクトルを含む知覚表現を受け取り、

ニューラルネットワークメモリユニットによって表される複数のスロットベクトルを初期化し、複数のスロットベクトルのそれぞれのスロットベクトルは、知覚表現に含まれる対応するエンティティを表すように構成され、

(i)キー関数によって変換された複数の特徴ベクトル、および(ii)クエリ関数によって

変換された複数のスロットベクトルの積に基づいてアテンション行列を決定し、アテンションマトリックスの複数の次元の各次元に沿った複数の値のそれぞれの各値は、各次元に沿った複数の値に関して正規化され、

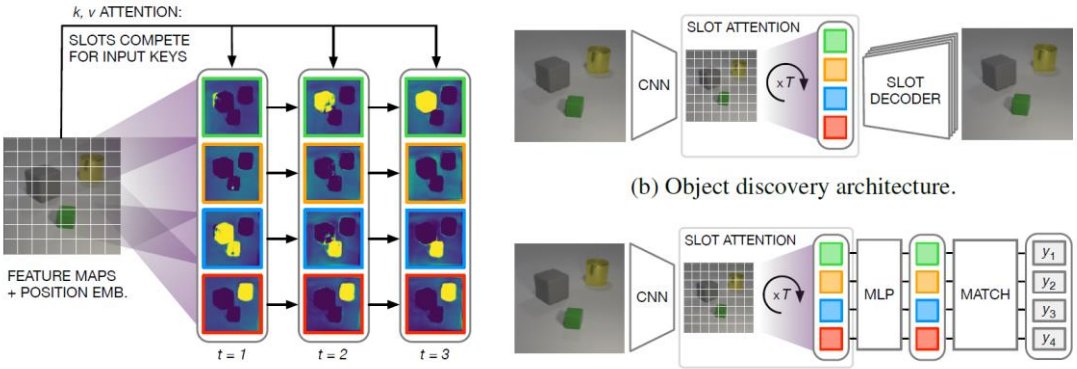
（i）バリュー関数によって変換された複数の特徴ベクトルおよび（ii）アテンション行列に基づいて更新行列を決定し、

ニューラルネットワークメモリユニットを介して、更新行列に基づいて複数のスロットベクトルを更新する。

4. 本特許に関する論文

本特許に関する論文“Object-Centric Learning with Slot Attention”が、Francesco Locatello 氏らにより公表されている[1]。

本論文では、畳み込みニューラルネットワークの出力などの知覚表現とインターフェイスし、スロットと呼ばれるタスク依存の抽象的な表現のセットを生成するアーキテクチャコンポーネントである SlotAttention モジュールを紹介している。



(a) Slot Attention module.

(b) Object discovery architecture.

(c) Set prediction architecture.

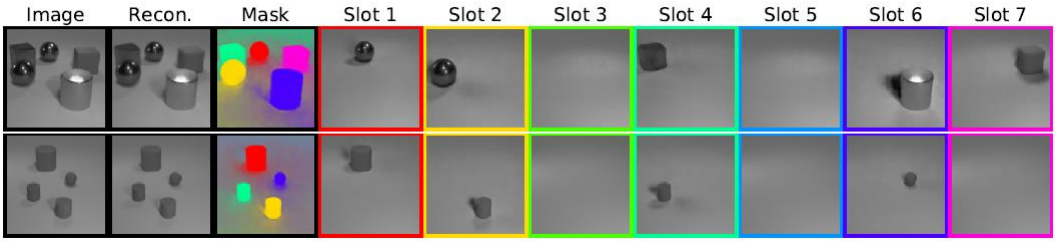
図 1

図 1 (a) は本特許で示されているスロットアテンションモジュールを示し、 (b)は本特許図 5A に示されている教師無しオブジェクト検出への適用例を示し、 (c)は、本特許図 5B に示されているラベル付けされたターゲット yi を使用した教師ありセット予測への適用例を示している。

[1] Francesco Locatello, DirkWeissenborn, Thomas Unterthiner, Aravindh Mahendran, Georg Heigold, Jakob Uszkoreit, Alexey Dosovitskiy, Thomas Kipf “Object-Centric Learning with Slot Attention” arXiv:2006.15055v2 [cs.LG] 14 Oct 2020

	CLEVR6	Multi-dSprites	Tetrominoes
Slot Attention	**98.8 ± 0.3**	**91.3 ± 0.3**	**99.5 ± 0.2***
IODINE [16]	**98.8 ± 0.0**	76.7 ± 5.6	**99.2 ± 0.4**
MONet [17]	96.2 ± 0.6	**90.4 ± 0.8**	—
Slot MLP	60.4 ± 6.6	60.3 ± 1.8	25.1 ± 34.3

上記テーブルは、マルチオブジェクトデータセットでの教師なしオブジェクト検出の調整済みランド指数（ARI：Adjusted Rand Index）スコア（%、5 シードの平均±標準偏差）を示している。Slot Attention のスコアの方が、他の従来のモジュールよりも高いことが理解できる。



上記図は、CLEVR6 のグレースケールバージョンでトレーニングされたスロットアテンションモデルのスロット毎の再構成とマスクを視覚化したものである。98：5±0：3%ARI を達成している。

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」、「ブロックチェーン 3.0(共著)」がある。