

AI 特許紹介(48)
AI 特許を学ぶ！究める！
～VQ-VAE 特許～

2023 年 1 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 DeepMind Technologies

出願日 2018 年 6 月 11 日

公開日 2020 年 6 月 11 日

公開番号 US20200184316

発明の名称 入力データ項目の離散潜在表現の生成

316 特許は、教師無し学習の一種である VAE(Variational AutoEncoder)を改良した VQ-VAE (Vector Quantised ベクトル量子化-VAE) に関し、エンコーダネットワークに、連続コードではなく離散コードを出力させ、また事前確率を静的ではなく学習させることで、VAE に散見される事後崩壊 (posterior collapse) を防止し、高品質の画像及び音声を生成するものである。

2.特許内容の説明

図 1A は、エンコーダシステム 100 およびデコーダシステム 150 を示す。

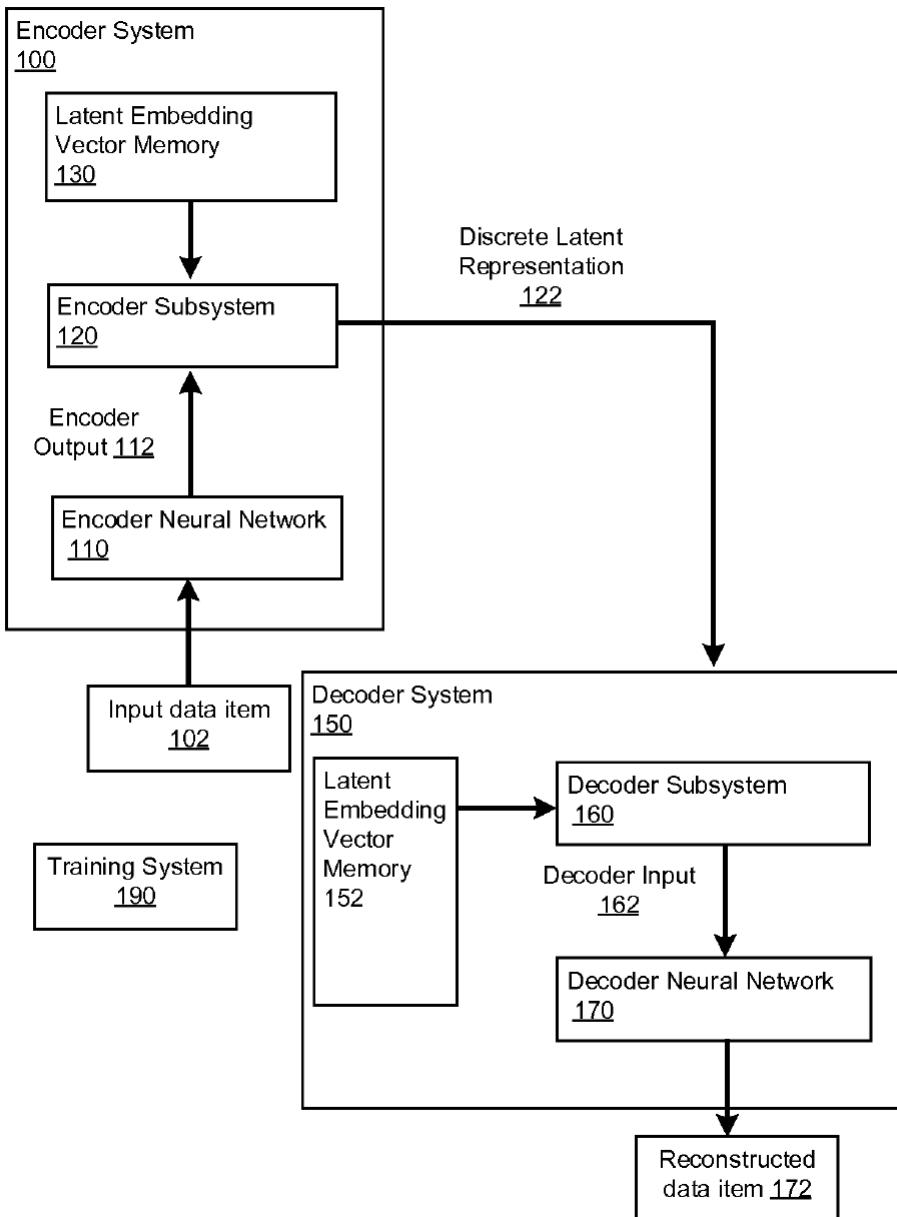


FIG. 1A

エンコーダシステム 100 は、入力データ項目 102 を受信し、入力データ項目 102 をエンコードして、入力データ項目 102 の離散潜在表現 122 を生成する。潜在表現 122 は、潜在空間における入力データ項目 102 の特徴を識別する数値表現である。デコーダシステム 150 は、入力データ項目 150 の離散潜在表現 122 を受け取り、入力データ項目 102 の再構成である再構成データ項目 172 を生成する。

エンコーダシステム 100 は、エンコーダニューラルネットワーク 110、エンコーダサブシステム 120、および潜在埋め込みベクトルのセットを格納する潜在埋め込みベクト

ルメモリ 130 を含む。

エンコーダニューラルネットワーク 110 は、入力データ項目 102 を処理して、一連のパラメータ（エンコーダネットワークパラメータ）に従って入力データ項目 102 のエンコーダ出力 112 を生成するようにトレーニングされたニューラルネットワークである。エンコーダ出力 112 は、1つまたは複数の潜在変数のそれぞれについて、それぞれのエンコードされたベクトルを含む。

入力データ項目 102 が画像である場合、各潜在変数は、二次元特徴マップ内の別個の空間位置に対応する。すなわち、エンコーダ出力 112 は、2次元特徴マップ内の各空間位置に対するそれぞれのエンコードされたベクトルを含み、各空間位置は異なる潜在変数に対応する。これらの場合、エンコーダニューラルネットワーク 110 は、画像を処理してエンコーダ出力を生成する畳み込みニューラルネットワークとすることができます。

入力データ項目 102 がオーディオデータ、例えば生のオーディオ波形である場合、各潜在変数は、シーケンス内の別個の位置に対応する。すなわち、エンコーダニューラルネットワーク 110 は、エンコードされたベクトルのシーケンスとしてエンコーダ出力 112 を生成し、シーケンス内の各位置は異なる潜在変数に対応する。これらの場合、エンコーダニューラルネットワーク 110 は、オーディオデータのシーケンスを受け取り、エンコードされたベクトルのシーケンスを生成する拡張畳み込みニューラルネットワークである。

連続潜在表現を生成するシステムとは異なり、エンコーダ出力を入力データ項目 102 の表現として使用する代わりに、エンコーダサブシステム 120 は、エンコーダ出力 112 およびメモリ 130 内の潜在埋め込みベクトルを使用して、入力データ項目 102 の離散潜在表現 122 を生成する。

特に、潜在変数ごとに、エンコーダサブシステム 120 は、メモリ 130 内の潜在埋め込みベクトルのセットから、潜在変数のエンコードされたベクトルに最も近い潜在埋め込みベクトルを決定する。例えば、サブシステム 120 は、潜在埋め込みベクトルのセットに対する最近傍ルックアップまたは任意の他の適切な距離メトリックを使用して、所与のエンコードされたベクトルに最も近い潜在埋め込みベクトルを決定する。次に、サブシステム 120 は、離散潜在表現 122 に、潜在変数ごとに、潜在変数のエンコードされたベクトルに最も近い潜在埋め込みベクトルを識別するデータを含める。

エンコーダシステム 120 は、離散潜在表現 122 を、デコーダシステム 150 が離散潜

在表現にアクセスできるように、コンピュータによってアクセス可能なローカルメモリに格納する。

デコーダシステム 150 は、デコーダサブシステム 160 およびデコーダニューラルネットワーク 170 を含む。デコーダサブシステム 160 は、離散潜在表現 122 を受信し、潜在埋め込みメモリ 152 内の潜在埋め込みベクトルを使用してデコーダ入力 162 を生成する。

潜在埋め込みメモリ 152 は、通常、潜在埋め込みメモリ 130 と同じ潜在埋め込みベクトルを格納する。デコーダ入力 162 は、潜在変数ごとに、離散潜在表現 122 内の潜在変数について識別される潜在埋め込みベクトルを含む。デコーダシステム 150 は、エンコーダシステム 160 と同じ潜在埋め込みベクトルにアクセスできるため、離散潜在表現 122 は、潜在埋め込みベクトル自体を含む必要はなく、代わりに、デコーダシステム 150 に知られている（すなわち解決される）潜在埋め込みベクトルの識別子を含む。

デコーダニューラルネットワーク 170 は、デコーダ入力 162 を処理して、パラメータセット（デコーダネットワークパラメータ）に従って入力データ項目 102 の再構成 172 を生成するように訓練されている。

デコーダニューラルネットワークがデコーダ入力から高品質の再構成を生成できるようにするために、トレーニングシステム 190 は、エンコーダニューラルネットワーク 110 とデコーダニューラルネットワーク 170 を共同でトレーニングして、エンコーダネットワークパラメータとデコーダネットワークパラメータのトレーニング値を決定する。また、メモリ 130（および 152）内の潜在埋め込みベクトルを調整して、潜在埋め込みベクトルが入力データ項目の特徴を効果的に表現できるようにする。

共同トレーニングの後、システム 100 は、より強力なエンコーダニューラルネットワークを採用することができ、システム 150 は、トレーニングで使用されたネットワークよりも強力なデコーダニューラルネットワークを採用して、推論時のシステムのパフォーマンスを向上させることができる。

図 1B は、入力データ項目 102 が画像である場合のシステム 100 および 150 の動作の一例を示す。

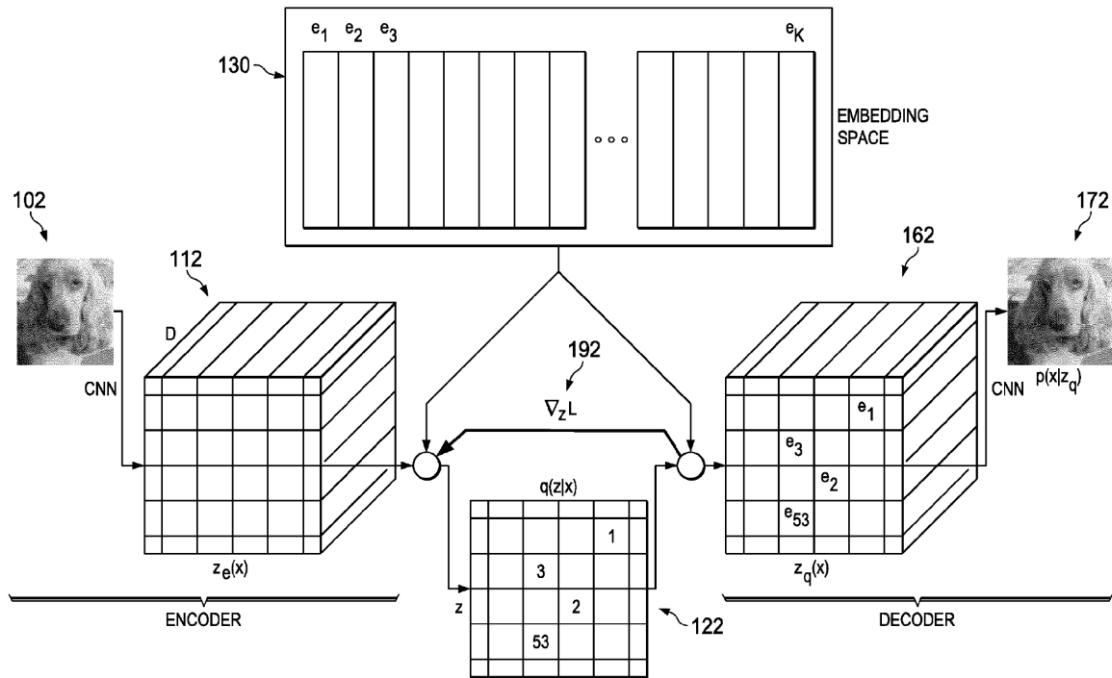


FIG. 1B

図 1B の例では、システム 100 は、入力データ項目（画像）102 を受信し、入力画像の再構成 172 を出力として生成する。エンコーダニューラルネットワーク 110 は、入力画像を処理してエンコーダ出力 112 を生成する。図 1B に示すように、エンコーダ出力 112 は、2 次元特徴マップ内の各空間位置に対するそれぞれの D 次元ベクトルであり、各空間位置はそれぞれの潜在変数に対応する。したがって、エンコーダ出力 112 は、複数の潜在変数のそれぞれについて、それぞれのエンコードされたベクトルを含む。

次に、システム 100 は、エンコーダ出力 112 およびメモリ 130 に格納された潜在埋め込みベクトルのセットを使用して離散潜在表現 122 を生成する。特に、図 1B の例では、メモリ 130 は、K 個の潜在埋め込みベクトル e_1 から e_K を格納する。

潜在表現 122 を生成するために、システム 100 は、潜在変数のそれぞれについて、潜在変数のエンコードされたベクトルに最も近い K 個の潜在埋め込みベクトルの潜在埋め込みベクトルを、例えば最近傍ルックアップを使用して識別する。

次いで、システム 100 は、潜在変数のそれぞれについて、潜在変数のエンコードされたベクトルに最も近い潜在埋め込みベクトルを識別する潜在表現 122 を生成する。図 1B に示すように、第 1 の空間位置に対応する潜在変数の場合、表現 122 は潜在埋め込みベクトル e_1 を識別し、第 2 の空間位置に対応する潜在変数の場合、表現は潜在埋め込みベクトル e_{53} を識別する。

次に、システム 150 は、潜在埋め込みベクトルおよび潜在表現 122 を使用してデコーダ入力 162 を生成する。システム 150 は、デコーダ入力 162 を、複数の空間位置のそれぞれにD次元ベクトルを有する三次元特徴マップとして生成する。任意の所与の空間位置におけるD次元ベクトルは、潜在表現 122 内の対応する潜在変数について識別された潜在埋め込みベクトルである。したがって、デコーダ入力 162 は、第 1 の空間位置に潜在埋め込みベクトル e1 を含み、第 2 の空間位置に潜在埋め込みベクトル e53 を含む。

次に、システム 150 は、デコーダニューラルネットワーク 170 を使用してデコーダ入力 162 を処理し、入力データ項目 102 の再構成 172、すなわち潜在表現 122 に基づく入力の推定値である入力画像と同じサイズの画像を生成する。

図 1B の例が、エンコーダ 110 およびデコーダ 150 のトレーニング中に生成される再構成を示す場合、トレーニングシステム 190 は、エンコーダネットワークパラメータの値を更新するために勾配をエンコーダに逆伝播する必要がある。トレーニングシステム 190 は、エンコーダネットワークパラメータを更新できるようにするために、勾配 192 をデコーダ入力からエンコーダ出力にコピーすることができる。

3. クレーム

316 特許のクレーム 1 は以下の通りである。

1. システムにおいて、

潜在的な埋め込みベクトルのセットを格納するためのメモリと、
1つまたは複数のコンピュータと、1つまたは複数のコンピュータによって実行されると、1つまたは複数のコンピュータに実行させる命令を格納する1つまたは複数の記憶装置とを備え、

エンコーダニューラルネットワークは以下のように構成され、

入力データ項目を受け取り、

1つ以上の潜在変数のそれぞれについて、それぞれのエンコードされたベクトルを含むエンコーダ出力を生成するために、入力データ項目を処理し、

サブシステムは以下のように構成され、

入力データ項目のエンコーダ出力を得るために、入力データ項目をエンコーダニューラルネットワークへの入力として提供し、

エンコーダ出力から入力データ項目の離散潜在表現を生成し、

潜在変数のそれぞれについて、メモリ内の潜在埋め込みベクトルのセットから、

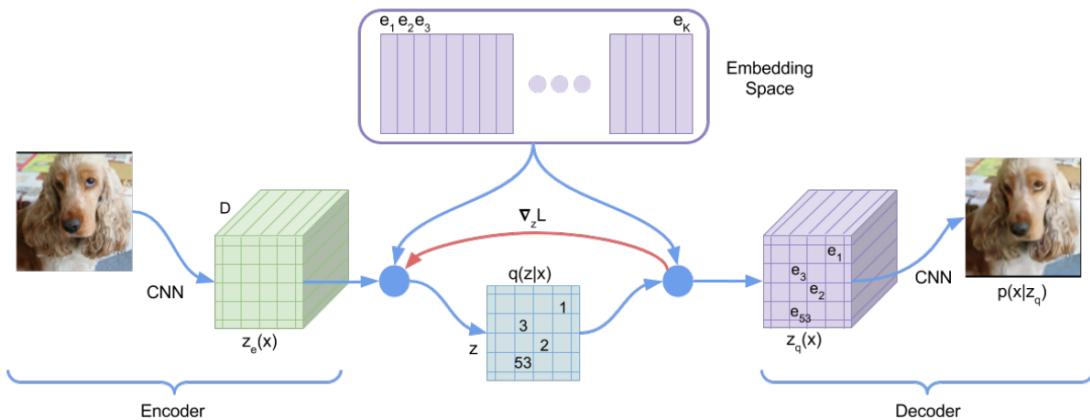
潜在変数のエンコードされたベクトルに最も近い潜在埋め込みベクトルを決定する。

4. 本特許に関連する論文

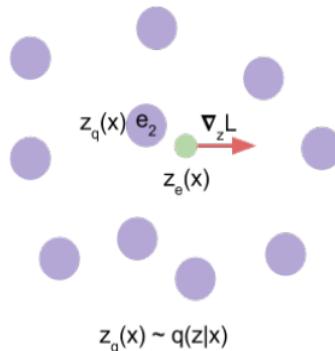
本特許に関する論文 “Neural Discrete Representation Learning”¹が、DeepMind の Aaron van den Oord 氏らにより公表されている。

本論文では離散表現を学習するシンプルで強力な生成モデル VQ-VAE が提案されている。VQ-VAE は従来の VAE と以下の点で相違する。エンコーダネットワークは、連続コードではなく離散コードを出力し、事前確率は静的ではなく学習される。

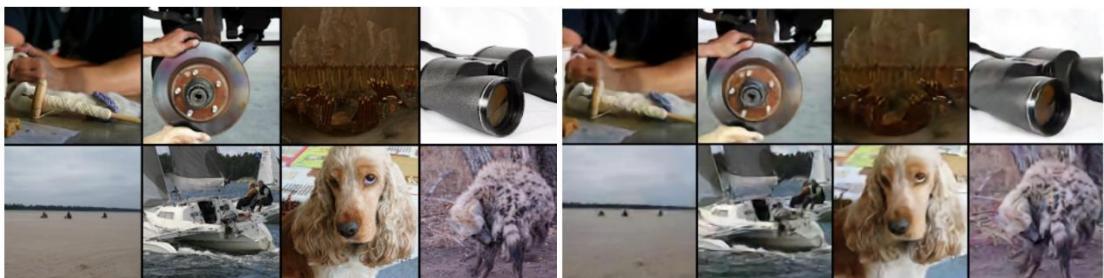
VQ-VAE は離散潜在表現を学習するために、ベクトル量子化 (VQ) のアイデアを取り入れている。これにより、モデルは VAE フレームワークで典型的にみられる「事後崩壊 posterior collapse」（強力な自己回帰デコーダと組み合わせた場合、潜在変数は無視される）の問題を回避できる。



¹ Aaron van den Oord, Oriol Vinyals, Koray Kavukcuoglu “Neural Discrete Representation Learning” arXiv:1711.00937v2 [cs.LG] 30 May 2018



上図の上側は VQ-VAE の構成図であり、動作内容は特許欄で説明した通りである。上図下側は、埋め込みスペースを可視化したものである。エンコーダ $z(x)$ の出力は、最も近い点 e_2 にマッピングされる。勾配 $\nabla_z L$ (赤) は、エンコーダをプッシュしてその出力を変更する。これにより、次のフォワードパスで構成が変更される。



上記左図は、ImageNet 128x128x3 の画像、右図は 32x32x1 潜在空間、K=512 の VQ-VAE からの再構成である。なお、K は離散潜在空間のサイズである。各画像とも正確に再構成されていることが理解できる。

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 3](#)」、「[ブロックチェーン 3.0](#)(共著)」がある。