

畳み込み変分オートエンコーダを用いた三次元構造物の認識と生成 Recognition and generation of 3D structure using convolutional variational autoencoder

立命館大学 ○ 植田考哉, 瀬尾昌孝, 西川郁子

T.Ueda, M.Seo and I.Nishikawa
Ritsumeikan University

Abstract Variational autoencoder (VAE) is a deep generative model with observation variables and latent variables. Encoder and decoder are given by the conditional distribution of these variables. The learned VAE can be used to generate the complete data by supplementing the missing part through the encode and decode operations on a partially missing observation data. In this study, recognition and generation of 3D structures were done by convolutional variational autoencoder (CVAE). We applied CVAE to the building construction of ordinary houses, and generated remaining parts from the partial solution. We also verified the feasibility of the generated data.

1 はじめに

深層生成モデルでは、観測データの分布自体を学習することにより、学習データから共通の特徴を抽出し、それを持つデータを生成できる。そこで本発表では、これを用いて最適化問題の求解を試みた。すなわち、組合せ最適化問題における良質な実行可能解を学習したモデルを用いて、部分解を入力し、実行可能解となるよう残りの部分を生成させる。適用事例として、3次元構造物である住宅構造設計を取り上げた。

一般住宅の住宅構造は、柱、大梁など様々な部材を組合せて配置することで得られる。耐震強度の各種制約条件の充足は必須であり、その上で効率的な配置を求める組合せ最適化問題である。強度制約はいずれも力学的な条件であり、部材の相互配置に依存する。それに対して本研究では、生成モデルを用いることで、一定の部材配置を与えた上で、制約条件を満たすように残りの部材配置を生成することを目指す。

2 変分オートエンコーダ

変分オートエンコーダ (Variational autoencoder: 以下VAE)は Kingma らによって提唱された、可視変数と隠れ変数からなる生成モデルである[1]。

隠れ変数 \mathbf{z} の分布 $p_{\theta}(\mathbf{z})$ からデコーダ $p_{\theta}(\mathbf{x}|\mathbf{z})$ によって可視変数 \mathbf{x} が生成されるとし、以下の周辺尤度を最大にすることを考える。

$$\log p_{\theta}(\mathbf{x}) = \log \int p_{\theta}(\mathbf{x}|\mathbf{z})p_{\theta}(\mathbf{z})d\mathbf{z} \quad (1)$$

ただし、 θ はモデルパラメータであり、隠れ変数の分布は多変量標準ガウス分布 $p_{\theta}(\mathbf{z}) = \mathcal{N}(\mathbf{z}|\mathbf{0}, \mathbf{I})$ と仮定する。

次に、可視変数から隠れ変数の分布を推論するためのエンコーダ $q_{\phi}(\mathbf{z}|\mathbf{x})$ (ϕ はモデルパラメータ)を導入し、多変量ガウス分布 $q_{\phi}(\mathbf{z}|\mathbf{x}) = \mathcal{N}(\mathbf{z}; \mu_{\phi}, \sigma_{\phi}^2 \mathbf{I})$ とする。式(1)の周辺尤度は以下の変分下界となり、この下界を最大化することが学習の目的になる。

$$\begin{aligned} \log p_{\theta}(\mathbf{x}) &\geq E_{q_{\phi}(\mathbf{z}|\mathbf{x})}[\log p_{\theta}(\mathbf{x}|\mathbf{z})] \\ &\quad - D_{KL}(q_{\phi}(\mathbf{z}|\mathbf{x})||p_{\theta}(\mathbf{z})) \\ &= \mathcal{L}(\mathbf{x}) \end{aligned} \quad (2)$$

式(2)の第2項は事前分布とエンコーダとの誤差であり最小にする。第1項の期待値はSGVB[2]と呼ばれる手法を用い以下のように近似的に導出し最大になるようにパラメータ $\{\phi, \theta\}$ を調節する(L はサンプリング回数、本実験では $L = 1$ とした)。

$$E_{q_{\phi}(\mathbf{z}|\mathbf{x})}[\log p_{\theta}(\mathbf{x}|\mathbf{z})] \cong \frac{1}{L} \sum_{l=1}^L \log p_{\theta}(\hat{\mathbf{x}}|\mathbf{z}^{(l)}) \quad (3)$$

$$\begin{aligned} \text{ただし} \mathbf{z}^{(l)} &= \mu(\mathbf{x}) + \varepsilon^{(l)} \odot \sigma(\mathbf{x}) \\ \varepsilon^{(l)} &\sim \mathcal{N}(\mathbf{0}, \mathbf{I}), \odot \text{は各要素の積} \end{aligned}$$

3 対象課題

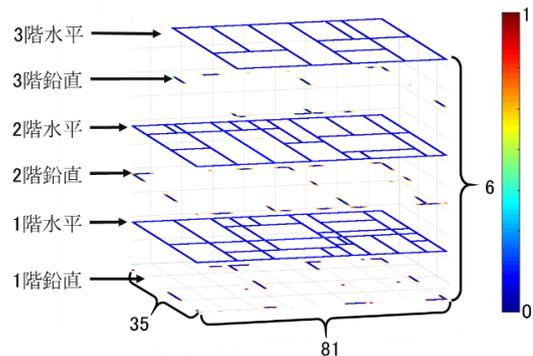


図 1: 対象とした住宅の部材配置例

部材配置の対象は図1に示す3階建の住宅とし、3次元ボクセルデータ($W35 \times D81 \times H6$)で表す。各ボクセル値は部材の体積占有率を示す(以下、部材量)。各階は、柱、耐力壁からなる鉛直面と、大梁、小梁、水平ブレースからなる水平面によって構築される。ここで、水平ブレースは、小梁配置によって決まるため以下では取り除いた(図1は除去後)。

4 データ生成実験

以下では、前節で示した住宅を対象に、小梁以外の部材を配置した部分解に対して、学習済みのVAEを用いて小梁を配置し、その実行可能性を検証する。

4.1 データセット

小梁を含めた全部材が配置されたボクセルデータを全部材データと呼ぶことにする。これらは構造計算により実行可能性が判定できる。まず、実行可能解 15,000 を用いて、2章に示した方法でVAEを学習した。次に、実行可能および不可能解を各 1,000 データ準備し、各データから小梁のみを取り除いたものを小梁なしデータと呼ぶ。

4.2 ネットワーク構造

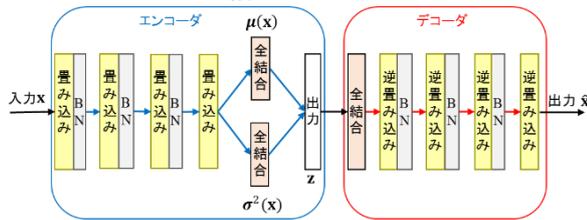


図 2: エンコーダ, デコーダのネットワーク構造

実験で用いるネットワーク構造を図 2 に示す。エンコーダは畳み込みを 4 層、デコーダは畳み込みのフォワード計算が逆方向となる逆畳み込みの 4 層から構成される。各層の活性化関数には *rectified linear unit* を使用し、*Batch Normalization*[2](BN)を経て活性化関数に入力する。また、最適化手法には *Adam*[3]を用いた。

エンコーダ $q_{\phi}(\mathbf{z}|\mathbf{x})$ は $\mathcal{N}(\mathbf{z}; \boldsymbol{\mu}_{\phi}, \boldsymbol{\sigma}_{\phi}^2 \mathbf{I})$ とし、最終層は混合ガウス分布の平均と分散を出力する。また、デコーダはベルヌーイ分布 $\mathcal{B}(\mathbf{x}|\boldsymbol{\pi}_{\theta}(\mathbf{z}))$ とし、パラメータ $\boldsymbol{\pi}_{\theta}$ は次式で与える (MLP は多層パーセプトロンの意)。

$$\boldsymbol{\pi}_{\theta}(\mathbf{z}) = \text{Sigmoid}(\text{MLP}_{\theta}(\mathbf{z})) \quad (4)$$

なお、ネットワークへのデータ入力は、高さ 6 のボクセルデータを、6 枚のピクセルデータとして入力した。

4.3 データ生成法

以下の手順で、学習済み VAE を用いて小梁を生成する。小梁なしデータ \mathbf{x} に対して、1~3 階の水平面から大梁が存在しないボクセルの集合を \mathbf{x}_s 、それ以外を \mathbf{x}_v とする。

- (1) \mathbf{x}_s の各要素に対して $[0,1]$ の一様乱数で初期化。初期値を $\mathbf{x}^{(k)} = \{\mathbf{x}_v^{(k)}, \mathbf{x}_s^{(k)}\} (k=0)$ と設定
- (2) $\mathbf{x}^{(k)}$ を学習済み VAE からエンコード→デコード操作によって $\hat{\mathbf{x}}^{(k)}$ を取得
- (3) $\mathbf{x}^{(k+1)} = \{\mathbf{x}_v^{(0)}, \hat{\mathbf{x}}_s^{(k)}\}$ とし $k = k + 1$ に更新
- (4) k が K 以下であれば(2)へ、それ以外は終了

VAE ではデコードしたデータに微量のノイズが含まれる。そこで全部材データの全データセットの最小値を閾値とし、それ以下の値はデコードによるノイズとみなして全て 0 にする。次に結果を示す実験では、 $K = 500$ とした。

4.4 生成結果

実行可能解から作成した小梁なしデータに対して、VAE が生成したデータを図 3 に例示する。

図 3(a)は元になる実行可能解、同図(b)は(a)から作成した小梁なしデータ、(c)は(b)に対して繰り返し回数 $K = 500$ で VAE が出力したデータ、(d)は(c)から前述の閾値でノイズを除去したデータである。得られた(d)は(a)に類似した小梁配置となっていることが分かる。他のデータに対しても同様の結果が得られた。

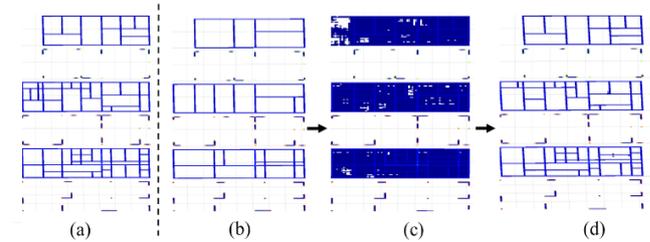


図 3: VAE による小梁の配置結果の例

4.5 生成されたデータの実行可能性の検証

VAE による小梁配置で得られた全部材データの実行可能性を判定する。ただし、VAE からの出力であるボクセルデータから直接構造計算は出来ないため、ここでは実行可能性を判別するネットワークを構築し、簡便な判定を行った。ネットワークには VGG ネット[4]を使用し、実行可能/不可能の 2 クラス分類を行う。データセットは 4.1 の全部材データと同様に、実行可能/不可能それぞれ 15000 データを学習用、各 1000 データをテスト用とした。

次に学習済み VGG による分類結果を示す。学習データでの正解率は 99.5%、テストデータでの正解率は 91.5% となった。これを用いて 4.3 の方法で VAE が生成したデータを分類した結果を表 1 に示す。小梁なしデータが実行可能解から作成されたものをグループ 1、不可能解から作成されたものをグループ 2 とした。グループ 1 の結果から、実行可能解の部分解からは高い割合で再び実行可能解を生成していること、グループ 2 の結果から、実行不可能解の部分解からでも一定の割合で実行可能解を生成できることがそれぞれ分かる。

表 1: VGG が実行可能と判定した VAE 生成データ数

グループ 1	グループ 2
891/1000	307/1000

5 まとめ

3次元構造の設計問題を対象に、深層オートエンコーダを用いて、部分解から実行可能解を生成する方法を提案した。今後は実行可能性のより厳密な検証や生成モデルの一般性を検討したい。

参考文献

- [1] Diederik P. Kingma, Max Welling, Auto-Encoding Variational Bayes, ICLR, 2014
- [2] S.Ioffe, C.Szegedy, Batch Normalization: Accelerating deep network training by reducing, ArXiv:1502.03167, 2015
- [3] Diederik P. Kingma, Jimmy Lei Ba, Adam: A method for stochastic optimization, ICLR, 2015
- [4] Karen Simonyan, Andrew Zisserman, Very deep convolutional networks for large-scale image recognition, ICLR, 2015