Vol.2018-MPS-118 No.35
Vol.2018-BIO-54 No.35
2018/6/15

深層学習を用いた組合せ最適化問題への強化学習の適用

三木 彰馬[1,a)]　山本 大輔[1]　榎原 博之[2,b)]

概要：本論文では代表的な組合せ最適化問題の 1 つである巡回セールスマン問題（TSP）に注目し，深層学習と強化学習を適用した解法を提案する．本手法では，畳み込みニューラルネットワークを用いて最適経路を画像として学習することで，最適経路に含まれうる辺の分布である優良エッジ分布を求め，これにより計算される辺の評価値である優良エッジ値を利用して近傍探索を行う．また，規模が大きい問題や他の組合せ最適化問題では最適解を求めることができない場合があるため，最適解ではなく最良解を用いた学習が必要である．そこで，強化学習を利用した学習方法についても検討する．これらの手法の性能を調べるために実験を行い，解の精度向上に関して有効性を検証する．

キーワード：組合せ最適化問題，巡回セールスマン問題，深層学習，強化学習，畳み込みニューラルネットワーク

Shoma Miki[1,a)]　Yamamoto Daisuke[1]　Hiroyuki Ebara[2,b)]

1. はじめに

組合せ最適化問題は計算機科学における基本的な問題の 1 つである．輸送や通信，製造，インフラ計画など，さまざまな分野における多くの課題は組合せ最適化問題として扱うことができ，現実社会での応用が期待されている．典型的な組合せ最適化問題の 1 つに巡回セールスマン問題（TSP : Traveling Salesman Problem）が挙げられる．TSP とは与えられたグラフにおいて，すべての頂点を 1 度だけ通るような巡回路のうちエッジ（辺）の距離の総和を最小とするものを求める問題である．とくに頂点が 2 次元平面上にあり，エッジの距離が頂点間のユークリッド距離として定義されるものを平面 TSP と呼ぶ．

組合せ最適化問題の解法はその解の精度の違いによって大きく分けられ，厳密解法やヒューリスティクスがある．厳密解法では列挙法や分枝限定法などを用いて最適解を求めるが，問題の規模が大きい場合には現実的な時間で最適解を求めることができない．一方，ヒューリスティクスでは解の精度が保証されないが短い時間で解を求められる可能性があり，その例として遺伝的アルゴリズム [1] などが挙げられる．これらの背景から，高速かつ高精度なヒューリスティクスの開発が重要視されている．

近年，機械学習および深層学習を用いた技術が活発に研究され，今までは困難であった課題を解決できる可能性がある手法として注目されている．深層学習では学習に長い時間を費やすことで問題の特徴量を事前に獲得し，高精度かつ高速な近似を実現する．このことから，組合せ最適化問題の解法において深層学習を利用することで，計算時間を削減しつつ，より高精度な解を求める解法を得ることが期待できる．

深層学習の技術が活発に応用されている領域の 1 つが画像処理である．畳み込みニューラルネットワーク（CNN : Convolutional Neural Network）は畳み込み演算を行う処理層で構成されるニューラルネットワークであり，画像認識をはじめ画像を入力とする問題において高い性能を示している．その例として Deep Convolutional Generative Adversarial Network（DCGAN）を用いた画像生成 [2], [3] などが挙げられる．

また，深層学習がさまざまな分野に適用されていくなかで，強化学習の重要性が増している．強化学習ではモデルの出力に対して報酬が与えられ，この報酬を最大化するように学習を行うことで，望ましい出力を獲得する．強化学習では教師データが不要であるため，教師データを用意することが難しく教師あり学習ができないような問題でも

[1] 関西大学大学院
Graduate School of Kansai University, Suita, Osaka, Japan
[2] 関西大学
Kansai University, Suita, Osaka, Japan
[a)] k154911@kansai-u.ac.jp
[b)] ebara@kansai-u.ac.jp

適用できる場合がある．深層学習を囲碁 AI に適用する手法 [4] では，盤面の情報を CNN に入力し，盤面上の各位置に対する評価値を近似する．その学習時には自己対戦によって生成した棋譜データを用いて強化学習を行い，CNN の性能を改善することで高い棋力を実現している．

TSP を含む組合せ最適化問題に対して深層学習を適用する手法はいくつか研究されている [5], [6], [7]．問題が持つグラフ構造を利用した手法 [7] では，グラフに従ってネットワークの処理層を結合し，隣接する頂点間で特徴量を伝播させることで，グラフ構造を効率よく反映した特徴量の抽出を図っている．しかしこの手法はすべての頂点が互いに隣接する TSP では精度の高い特徴抽出が難しく，特徴量を伝播する近傍を制限するなどの工夫が必要である．また，組合せ最適化問題では計算コストの面から最適解を求めることが難しい場合があるため，このような問題では最適解ではない解を用いて学習を行う必要がある．これに関して Bello らによる研究 [6] では，LSTM（Long Short-Term Memory）を用いたモデルに対して解の目的関数を報酬として強化学習を行うことで，最適解を教師データとして必要としない学習方法を提案している．

本論文では平面 TSP に注目し，CNN を用いてエッジの評価値を計算する手法と，距離の代わりにエッジの評価値を用いるヒューリスティクスを提案する．この手法では TSP とその解を画像として扱い，CNN により最適経路の画像を近似した優良エッジ分布を計算，その出力に従ってエッジを選ぶことで解を求める．また，最適解を必要としない学習を実現するため，優良エッジ分布に強化学習を適用する方法を検討する．これらの手法の性能を検証するために実験を行い，その有効性について議論する．

2. CNN を用いたエッジの評価

平面 TSP では各頂点が 2 次元ユークリッド座標で与えられ，点および線の描画によってその問題と解を画像として表現できる．ここである平面 TSP の問題例について，すべての頂点を描画した頂点画像 $N(x,y)$ と，その最適経路を描画した最適経路画像 $t(x,y)$ を定義する．ただしこれらの画素数を (S_1,S_2) とし，画像内の任意の画素の位置 (x,y) は $(x,y) \in \{1\,..\,S_1\} \times \{1\,..\,S_2\}$ を満たす．また，描画前の画素値を 0 とし，最大画素値 1 で点および線を描画する．

頂点画像 $N(x,y)$ を入力したときその問題の最適経路画像 $t(x,y)$ を出力するようなモデルを考える．このようなモデルが得られれば，その出力画像に表されたエッジを選ぶことで最適経路を求めることができる．提案手法では CNN を用いることでこのモデルを近似し，その出力 $p(x,y)$ を優良エッジ分布（Good-Edge Distribution）と呼ぶ．優良エッジ分布の学習方法とこれを用いた解法の概要を**図 1** に示す．

優良エッジ分布はその出力 $p(x,y)$ を教師信号 $t(x,y)$ に近づけることを目的として学習を行う．1 つの問題例について式 (1) のように 2 乗誤差で表される損失関数を与え，これを最小化するように勾配降下法および誤差逆伝播法を用いて CNN の重みを更新する．

$$\text{loss} = \sum_{x=1}^{S_1}\sum_{y=1}^{S_2}(t(x,y)-p(x,y))^2 \tag{1}$$

学習により優良エッジ分布が得られたとき，各エッジ上における優良エッジ分布の平均を求めることで，そのエッジが最適経路に含まれる尤度を計算する．これを優良エッジ値（Good-Edge Value）と呼ぶ．

3. 優良エッジ値を用いたヒューリスティクス

優良エッジ値は各エッジが最適解に含まれる尤度であり，優良エッジ値が大きいエッジを選ぶことにより最適経路に近い解が得られると考えられる．そこで，優良エッジ分布を距離の代わりに使用するヒューリスティクスとして，貪欲法を応用した EV-greedy 法，2-opt 法を応用した EV-2opt 法を提案する．

3.1 EV-greedy 法

TSP における貪欲法（greedy algorithm）の 1 つとして，短いエッジを順番に巡回路に追加していくことで解を構築する手法が挙げられる．これに対して，EV-greedy 法では優良エッジ値が高い辺を巡回路に追加していくことで解を構築する．

3.2 EV-2opt 法

従来の 2-opt 探索法は TSP における近傍探索法の 1 つであり，2 つのエッジをつなぎかえ，経路の総距離が短くなる組合せを選択することで探索を行う．一方，EV-2opt 法では経路の優良エッジ値の総和が大きくなる近傍解を選択する．近傍探索法の遷移の方法には即時移動戦略と最良移動戦略があるが，実験時には即時移動戦略を採用する．

EV-greedy 法および EV-2opt 法ではエッジが交差しないことを保証できないため，これらを使って求めた解に対して距離を用いる通常の 2-opt 探索法を最後に適用することでさらに解の精度を改善できることが期待される．これらの手法を便宜上 EV-greedy+2opt，EV-2opt+2opt と表記する．

4. 強化学習の適用方法

これまでに述べた優良エッジ分布の学習方法は最適経路の画像を教師データとして用いる教師あり学習であった．しかしながら，大規模な問題例や他の組合せ最適化問題では現実的な時間で最適解を求めることができないといった理由から，十分な数の教師データを用意することが困難な場合がある．このような問題に対応するためには最適解を

Vol.2018-MPS-118 No.35
Vol.2018-BIO-54 No.35
2018/6/15

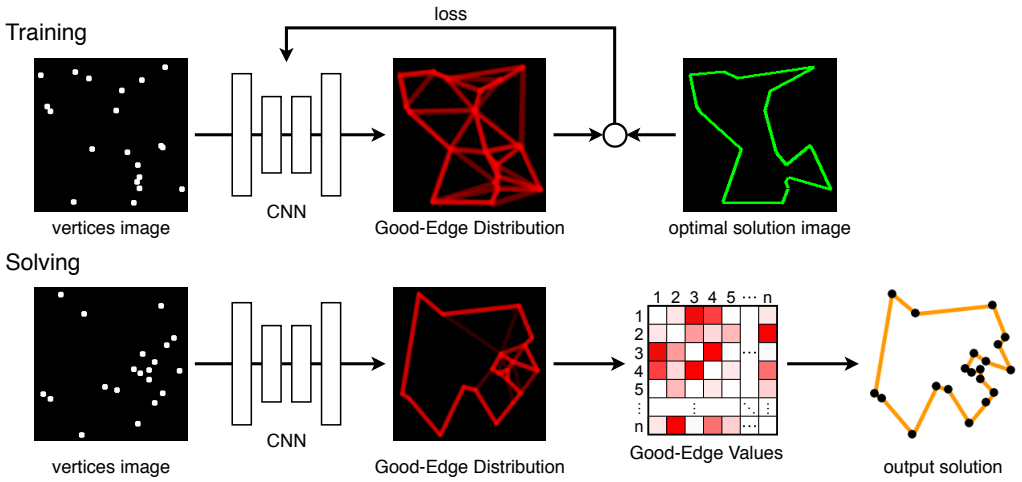


図 1　優良エッジ分布の学習方法に対する教師あり学習および解法の概要

Fig. 1　Overview of supervised learning and solving for the Good-Edge Distribution.

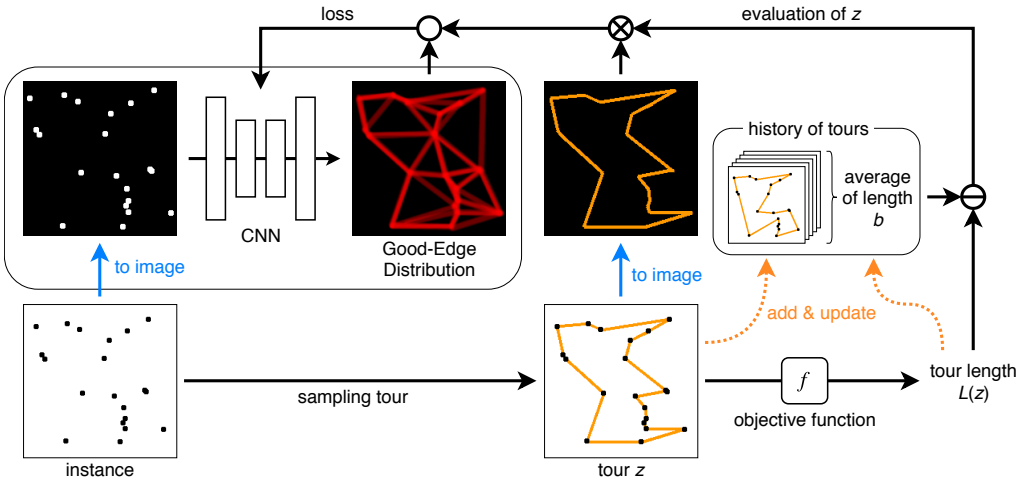


図 2　優良エッジ分布に対する強化学習の概要

Fig. 2　Overview of reinforcement learning for the Good-Edge Distribution.

必要としない学習アルゴリズムが必要であり，本研究では優良エッジ分布を求める CNN の学習方法としてを強化学習を適用する手法を提案する．

提案する学習方法の概要を図 2 に示す．TSP の目的関数および解の精度は巡回路の経路長として定義され，経路長が短いほど解の精度が良い．そこで巡回路の経路長が短いほど，その画像を出力しやすくなるように CNN を学習する．学習過程では学習用の任意の問題例に対してランダムに解をサンプリングし，その解の精度の良悪を評価する．解の評価の際には過去に生成した解の経路長の平均を基準値とし，経路長が平均よりも短ければその経路の出力を強化，長ければ弱化するように CNN を更新する．CNN の更新には，1 つの問題例と解 z に対して式 (2) に示すような損失関数として誤差逆伝播を行う．ただし，過去の解の経路長の平均を b とし，解 z の経路長を $L(z)$，経路を描画した画像の画素値を $t(z,x,y)$ とする．

$$\text{loss} = (L(z) - b) \sum_{x=1}^{S_1} \sum_{y=1}^{S_2} p(x,y)\, t(z,x,y) \qquad (2)$$

このように解をサンプリングし CNN を更新するまでの処理を 1 ステップとし，これを繰り返すことで学習を行う．また，1 ステップの処理が終わったとき，図中の破線矢印に示すようにサンプリングした解を過去の履歴として記録し、その経路長をもって指数移動平均により平均経路長 b の値を更新する．

ランダムに解をサンプリングするとき，完全にランダムな解を生成する手法では高精度な解をなかなか得ることができないため，学習がうまく進まない．そこで，学習途中の優良エッジ分布の性能を反映しつつ解をサンプリングする方法として，EV-ε-greedy 法を提案する．この手法では，EV-greedy 法において次に追加するエッジを選択する際，確率 ε でランダムなエッジを選ぶ．

5. 評価実験

本節では，提案手法の性能を評価するために実施した計算機実験の方法とその結果について説明する．

実験プログラムのプログラミング言語として Python を，機械学習用のライブラリとして TensorFlow を使用する．

5.1 実験に用いたデータ

学習に用いる平面 TSP の問題例として，頂点数 20〜100 の問題例を 20 万個，各頂点の座標を一様乱数により設定することで生成し，TSP ソルバー Concorde [8] を用いてその最適解を求める．教師信号および学習する CNN の入出力の画素数は $(S_1, S_2) = (192, 192)$ とし，頂点の描画半径を 1.5，エッジの描画幅を 1 としてそれぞれの問題例の入力信号 $N(x, y)$ と教師信号 $t(x, y)$ を作成する．また，テスト用問題例には訓練データと同様の手法で生成したランダム問題と，TSPLIB [9] に含まれる問題例から抜粋して使用する．

実験時の解の評価には最適解に対する経路長の誤差率を使用し，最適解の経路長 L^* に対して，経路長 L の解の誤差率 ε を式 (3) に従って計算する．

$$\varepsilon = \frac{L - L^*}{L^*} \times 100\ (\%) \tag{3}$$

5.2 教師あり学習による優良エッジ分布の学習

訓練データの問題例を用いて，教師あり学習による優良エッジ分布の学習を行い，EV-greedy+2opt および EV-2opt+2opt の性能について検証を行う．

優良エッジ分布のモデルには，**図 3** に示すような 14 層の畳み込み層と 4 層の転置畳み込み層からなる CNN を使用する．この CNN は U-Net [2], [10], [11] と類似の構造を持ち，画像の縮小を行うエンコーダと拡大を行うデコーダの各層が skip connection によって結合されている．出力層以外の層ではスロープ係数 0.2 の Leaky ReLU [12] を，出力層では恒等写像を活性化関数として用いる．学習時のミニバッチ数は 32 とし，学習アルゴリズムには Adam 法 [13] を使用する．

解法の性能評価では，テスト用問題例について各解法により得られた解の誤差率を比較する．EV-2opt+2opt は通常の 2-opt 法と同様に解の精度と計算速度が初期解に大きく依存するため，距離を用いた通常の貪欲法によって求めた解を初期解とすることで精度の安定と向上を図っている．また，TSPLIB の問題例に関しては，S2V-DQN（Stracture2Vec Deep Q-Learning）による実験結果 [7] も同様に比較する．

学習によって得られた優良エッジ分布と，提案した解法によって得られた解の出力例を**図 4** に示す．各図の (a) は与えた問題例の最適経路を，(b) はその問題に対する優良エッジ分布の出力を表す．(c) から (e) は EV-2opt+2opt の処理過程における解の例を示している．

図 4(b) より，優良エッジ分布が最適経路に含まれるエッジを多く持ち，最適経路の概形を表現できていることがわかる．ただし画像上で頂点が密集するような場所では，最適ではないエッジ上でも大きな値を出力することが多くみられた．この傾向は解を求める際に精度を下げる要因となるため，より大きな頂点数の問題例を訓練データとして使用することや，データ拡張などを行うことによって CNN の精度を改善する必要がある，

テスト用の問題例に対して解を求めたときの平均誤差率を**表 1** に示す．ただしここでは TSPLIB より頂点数 200 以下の問題例を 15 個を抜粋しており，問題例の名前に含まれる数字は頂点数を表している．この表より，誤差率の平均値は EV-2opt+2opt が最も低く，次いで EV-greedy+2opt となっており，貪欲法や 2-opt 法，S2V-DQN を下回った．また各問題例ごとに EV-2opt+2opt と既存手法を比較した場合でも，ほとんどの問題において平均誤差率は低い値となった．このことから，提案手法を用いることによる解の精度向上を確認することができた．

5.3 強化学習の適用

強化学習を用いて優良エッジ分布を学習する手法を実装し，その効果を確認する．強化学習は長い学習時間が必要であるため，今回は実験時間の都合上，教師あり学習を行った後に強化学習を行い，性能の変化をみることで効果を検証する．訓練データには教師あり学習の際に使用したものと同じデータセットを使用し，学習時には最適解を使用せず，頂点座標の情報のみを訓練データとして与える．解を求めるアルゴリズムには，学習中のサンプリング時には EV-ε-greedy を，テスト時には EV-greedy を使用し，学習による効果を調べるためにテスト用の問題例に対する相対誤差を求め，その推移を観測する．ただし，テスト用問題にはランダムな問題例 11 個と TSPLIB の問題例 15 個を使用する．

強化学習の学習過程における相対誤差の変化を**図 5**，優良エッジ分布の出力例を**図 6** に示す．ただし図 5 中の相対誤差はすべてのテスト用問題例に対する相対誤差の平均値を表している．図 5 より，強化学習を開始してから 1500 ステップ付近までは解の相対誤差が低下しており，強化学習による精度の改善がみられる．しかし，それ以降は相対誤差が上昇しており，解が平均的に悪化している．また，図 6 の優良エッジ分布の出力をみた場合，ステップを重ねるにつれてより最適解の形が強化され，不要なエッジが除去されていく様子がうかがえる．しかし 3000 ステップでは一部の頂点の周囲で優良エッジ分布の出力が無くなるという現象がみられ，4500 ステップではすべての出力が

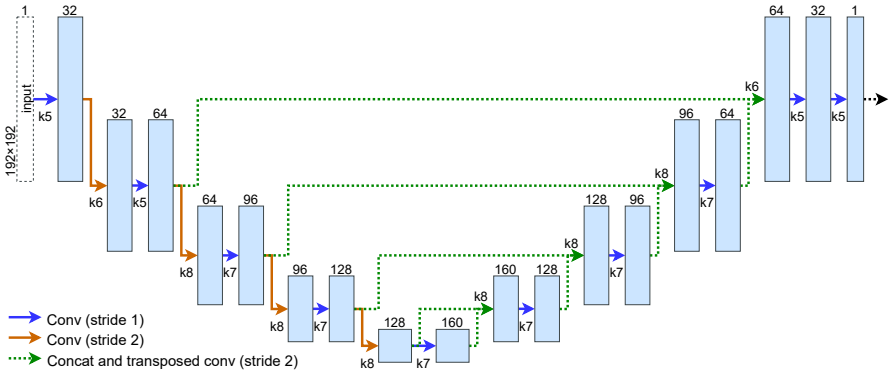
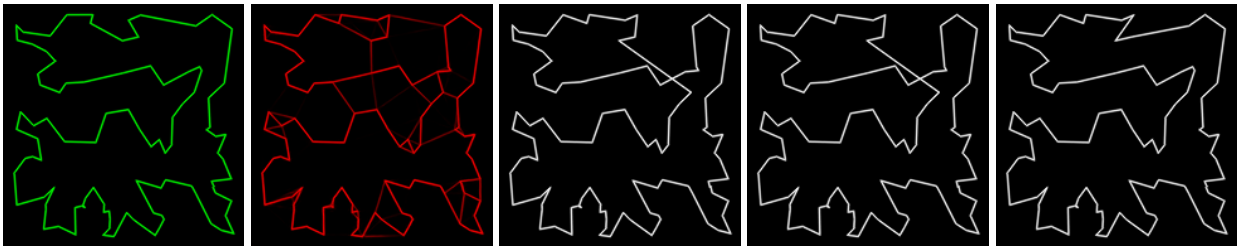


図 3　CNN の構造
Fig. 3　Structure of CNN.



(a) 最適経路
(a) Optimal tour.

(b) 優良エッジ分布
(b) Good-Edge Distribution.

(c) 貪欲法による初期解（誤差率 16.979%）
(c) Greedy ($\varepsilon = 16.979\%$).

(d) EV-2opt（誤差率 2.996%）
(d) EV-2opt ($\varepsilon = 2.996\%$).

(e) EV-2opt+2opt（誤差率 0.721%）
(e) EV-2opt+2opt ($\varepsilon = 0.721\%$).

図 4　優良エッジ分布と解の出力例（rd100）
Fig. 4　Output of the Good-Edge Distribution and solutions (rd100).

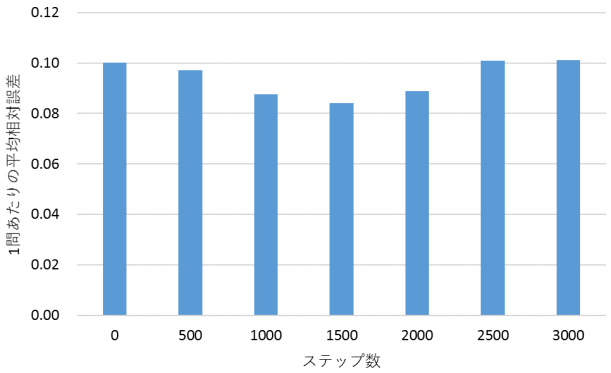


図 5　強化学習による解の相対誤差の推移
Fig. 5　Relative errors during reinforcement learning.

消えてしまっている．このように最適なエッジの出力がなくなった場合，EV-greedy や EV-ε-greedy ではランダムにエッジを選んでしまうため，相対誤差が悪化にしたと思われる．このように学習が破綻してしまう問題を解決するために，Experience Replay などの学習を安定化する手法の導入や，学習アルゴリズムの再検討が必要であると考えられる．

6.　結論

本稿では平面 TSP に対して CNN を用いた解法と，このモデルの学習方法として強化学習を適用する手法を提案し，これらの性能を評価するために実験を行った．まず CNN を用いた解法では，CNN を用いて最適経路画像を近似した優良エッジ分布を算出し，ここから求められる優良エッジ値を距離に代わる指標として利用することで解の探索を行う．実験により，従来手法よりも解の誤差率が改善されたことを確認した．次に，最適解を教師データとして必要としない強化学習の適用方法について検討し，実験では教師あり学習のあとに強化学習を行うことで解の精度を改善できる可能性を示した．しかしながら，今回実験を行った強化学習では学習が不安定になる場合がみられたため，学習アルゴリズムの再検討が必要である．

謝辞　本研究の一部は，JSPS 科研費 18K11484 と，JSPS 科研費 17K01309，関西大学大学院理工学研究科高度化推進研究費，関西大学先端科学技術推進機構「緊急救命避難支援のための情報通信技術に関する研究開発」研究グループの助成を受けている．

参考文献

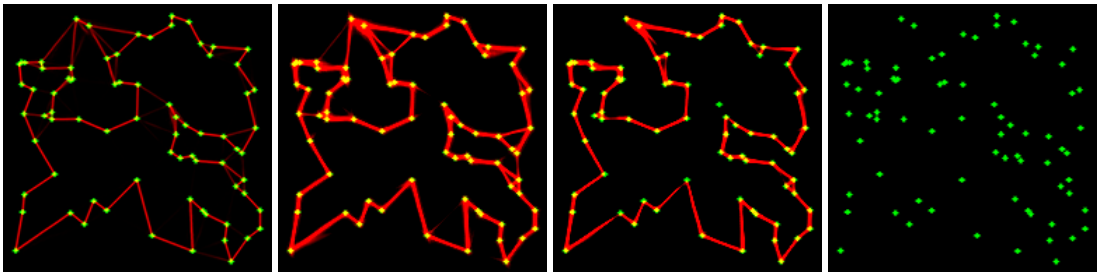
[1]　Nagata, Y. and Kobayashi, S.: A Powerful Genetic Algorithm Using Edge Assembly Crossover for the Traveling Salesman Problem, *INFORMS Journal on Computing*, Vol. 25, No. 2, pp. 346–363 (online), DOI: 10.1287/ijoc.1120.0506 (2013).
[2]　Isola, P., Zhu, J.-Y., Zhou, T. and Efros, A. A.: Image-

Vol.2018-MPS-118 No.35
Vol.2018-BIO-54 No.35
2018/6/15

表 1　各アルゴリズムの平均誤差率（TSPLIB）

Table 1　Average error ratios for each algorithms (TSPLIB).

問題例	**EV-greedy**	**EV-greedy+2opt**	**EV-2opt**	**EV-2opt+2opt**	greedy	2opt	S2V-DQN
eil51	6.338	**1.954**	11.162	2.958	24.648	5.070	3.052
berlin52	0.146	**0.000**	0.350	**0.000**	31.941	9.344	**0.000**
st70	**0.119**	0.415	1.489	0.452	11.111	5.963	3.111
eil76	**0.000**	**0.000**	0.372	0.037	8.736	6.952	4.833
pr76	7.766	1.319	6.648	2.188	36.370	3.750	**0.265**
rat99	6.639	**1.497**	5.925	1.936	18.910	7.605	5.698
kroA100	32.361	1.559	1.415	**0.477**	14.120	4.933	2.890
kroC100	11.302	1.588	4.170	**0.104**	12.270	6.255	1.566
rd100	17.649	2.027	2.996	**1.320**	16.979	6.414	3.148
eil101	10.016	**1.093**	5.469	1.932	24.483	6.677	4.769
lin105	2.031	0.852	5.709	**0.807**	16.601	5.639	4.479
bier127	11.912	3.022	8.553	**2.681**	19.493	6.478	2.785
ch130	14.599	3.279	9.814	3.205	18.216	7.117	**2.619**
kroA150	18.878	**2.418**	10.163	2.672	20.238	6.498	5.143
kroA200	12.982	4.342	7.388	**2.747**	17.659	6.396	5.438
総平均	10.182	1.691	5.442	**1.568**	19.452	6.339	3.320



(a) 0 ステップ　(b) 1000 ステップ　(c) 3000 ステップ　(d) 4500 ステップ

(a) 0 step.　(b) 1000 step.　(d) 3000 step.　(e) 4500 step.

図 6　強化学習による優良エッジ分布の学習過程

Fig. 6　Progress of reinforcement learning for Good-Edge Distribution.

to-Image Translation with Conditional Adversarial Networks, In *CVPR 2017*, *arXiv:1611.07004* (2016).

[3] Radford, A., Metz, L. and Chintala, S.: Unsupervised Representation Learning with Deep Convolutional Generative Adversarial Networks, *arXiv preprint arXiv:1511.06434* (2015).

[4] Silver, D., Huang, A., Maddison, C. J. et al.: Mastering the game of Go with deep neural networks and tree search, *Nature*, Vol. 529, pp. 484–489 (online), available from ⟨http://dx.doi.org/10.1038/nature16961⟩ (2016).

[5] Vinyals, O., Fortunato, M. and Jaitly, N.: Pointer Networks, *arXiv preprint arXiv:1506.03134* (2015).

[6] Bello, I., Pham, H., Le, Q. V., Norouzi, M. and Bengio, S.: Neural Combinatorial Optimization with Reinforcement Learning, *arXiv preprint arXiv:1611.09940* (2016).

[7] Dai, H., Khalil, E. B., Zhang, Y., Dilkina, B. and Song, L.: Learning Combinatorial Optimization Algorithms over Graphs, *arXiv preprint arXiv:1704.01665* (2017).

[8] Applegate, D. L., Bixby, R. E., Chvatal, V. and Cook, W. J.: Concorde TSP Solver, http://www.math.uwaterloo.ca/tsp/concorde/ (2006). Accessed 2018/2/22.

[9] Reinelt, G.: TSPLIB—A Traveling Salesman Problem Library, *ORSA Journal on Computing*, Vol. 3, No. 4, pp. 376–384 (online), DOI: 10.1287/ijoc.3.4.376 (1991).

[10] Ronneberger, O., Fischer, P. and Brox, T.: U-Net: Convolutional Networks for Biomedical Image Segmentation, *arXiv preprint arXiv:1505.04597* (2015).

[11] Yonetsuji, T.: PaintsChainer, https://github.com/pfnet/PaintsChainer (2017). Accessed 2018/2/22.

[12] Maas, A. L., Hannun, A. Y. and Ng, A. Y.: Rectifier nonlinearities improve neural network acoustic models, In *ICML Workshop on Deep Learning for Audio, Speech and Language Processing* (2013).

[13] Kingma, D. P. and Ba, J.: Adam: A Method for Stochastic Optimization, *arXiv preprint arXiv:1412.6980* (2014).