

Deep Learning を用いた街並み画像の分類と感性評価の推定

○高橋 秀彬*¹ 山田 悟史*²

キーワード：Deep Learning, Machine Learning, 感性・印象, 感性・印象評価の推定, 訪問意欲, 街並み, 景観

1. はじめに

近年、人工知能が専門家からビジネスマンまで急速に社会的関心を集めるようになった。またこれに伴い Deep Learning・Machine Learning という用語も広く知られるようになった。Deep Learning は人工知能を支える要素技術の一つであり、Machine Learning の一種である。発展途上の技術であるが、様々な分野で応用の検討が試みられている。画像の分類を例に Deep Learning の特徴を挙げれば、画像の特徴量を人が指定せずに、予め用意した分類に近づくように特徴量自体を探索させることができる点が挙げられる。これは同時に課題でもあるのだが、より端的に表現すれば、自動的にデータから機械に特徴量を学習させることが可能な手法と言える。この学習方法には様々な方法が提案されている。乱暴ではあるが学習方法について汎用的に言及すれば、分析者の意図する判断に関連付けたデータを大量に読み込ませ、判断条件を学習する何層ものニューラルネットワークを判断に近づくよう学習しながら繰り返し形成する、というものである。また教師無し分類も提案されている。

上述のような機械的な分類は、Deep Learning が着目される以前の Machine Learning においても提案されていたが、特徴量自体を探索する点に大きな特徴がある。この特徴を活かし、対象の特徴量記述・対象に対する判断が曖昧であった「画像認識」「音声認識」「自然言語処理」などの分野において特に適用が試みられている。これらの分野において Deep Learning が広まった契機の一つは、世界的な統計競技サイト「Kaggle」である。画像分類においてはモデルや推測の精度を競い合っており、Deep Learning を使った事例が以前の成果を大きく上回ったことが大きな影響を及ぼし、以降では Deep Learning を使った事例が増えている。音声認識についても 2011 年に Deep Learning を利用した方法が、従来の方法に比べエラー率が 20～30 ポイントも低いことが示された。このような技術的な報告だけでなく既に実装された事例もある。

以上のような Deep Learning を用いた人工知能技術は、建築分野への適用可能性も様々考えられ、やや過剰な印象もあるが期待感が高まっている。本研究が着目しているのは、感性評価・印象評価の学習と推定である。画像化した建築・都市デザインに対する感性・印象の推定については多くの研究が行われており、筆者らも一定の推定が可能であることを示してきた^{1) 2)}。これらの研究において特徴量は分析者が用意していた。一方で人の感性・印象は曖昧かつ複雑であり線形・非線形を問わずにモデル化が困難な性質がある。この

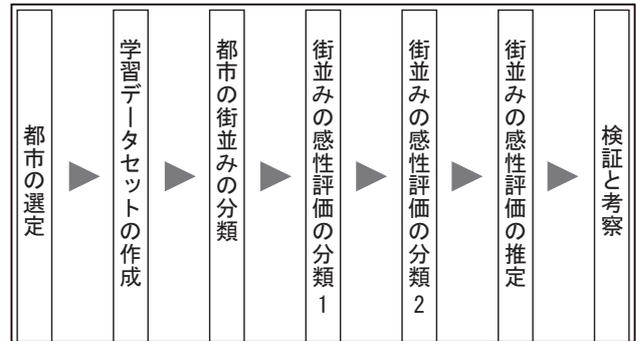


図1 研究の流れ

ような特徴は、既に Deep Learning において成果を挙げている分野と類似の性質であると考えられる。建築分野における応用の可能性の一つは、極論ではあるが人々や特定の建築家のデザイン感にもとづくデザイン生成である。しかし、これは人々や特定の建築家のデザイン感を仮想的に保存・再現できることが前提となる。そこで本研究では、建築・都市に関する画像に対する人の感性・印象評価を学習し推定するモデル構築に対する Deep Learning の適用可能性を検証する。

2. 研究概要

2.1 研究対象とデータセット

本研究の画像分類は画像認識でもあるが、画像認識の代表例である顔認識とは異なり、好きか嫌いか、といった感性や印象による分類である。顔などの特定に物体の画像認識として比較して曖昧な分類であるため、Deep Learning による適用可能性が有効であると考えた。本研究は「街並み」を対象に「都市の分類」と「感性評価」を学習・推定出来るかを検証した。図1に研究の流れを示す。

画像のデータセットの対象地は大都市や観光地^{注1) 注2)}から表1に示す21カ所を選定した。画像は、Google Earth のストリートビューから各箇所100枚、計2100枚を収集^{注3)}した。図2に21カ所中4ヶ所の対象街路と画像例を示す。

2.2 学習モデルについて

今回用いた学習モデルは Chainer ベースの GUI 環境で実行が可能な Network In Network³⁾ という深層ニューラル・ネットワークである^{注4)}。

3. 結果

3.1 街並みの都市の分類

都市の街並みの分類は Deep Learning による建築画

表 1 対象街路の一覧

ラベル	国	都市	通り
1	アメリカ	ニューヨーク	ウォール街通り
2	カナダ	セントジョンズ	キャボット通り
3	メキシコ	グアナファト	ポシトス通り
4	ペルー	クスコ	クエスタ・デ・サンタ・アナ
5	イギリス	ロンドン	オックスフォード通り
6	スコットランド	エディンバラ	コックバーン通り
7	スペイン	バルセロナ	カタルーニャ広場～ラセウ
8	ロシア	モスクワ	トゥヴェルスカヤ通り
9	イタリア	フィレンツェ	カルツァイウオリ通り
10	チェコ	プラハ	モステツカー
11	ポルトガル	ポルト	サン・ジョアン通り
12	南アフリカ	ケープタウン	ドーブ通り
13	アラブ	ドバイ	アル・ワーウッド通り
14	タイ	バンコク	ブレン・ナラー通り
15	香港	旺角	山東街
16	韓国	ソウル	北村韓屋村
17	台湾	九份	特定の名称なし
18	日本	東京	竹下通り
19	日本	大阪	道頓堀
20	日本	京都	産寧坂
21	オーストラリア	メルボルン	ホージャー～ラトリッジ・レーン

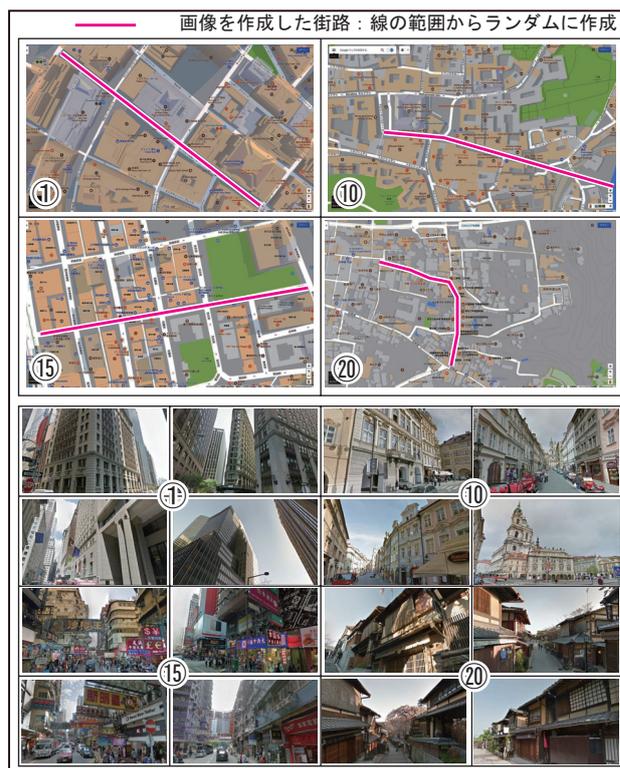


図 2 対象街路の範囲と作成した画像の一例

像の分類が可能であるかの基礎分析である。データセットの街並みを 21 種類都市からなるクラスに分類できるを検証した。結果を図 3^{注5)} に示す。グラフのオレンジ線・青線は学習時の正確性を表し、赤線・緑線はテスト時の正確性を表している。なお各画像の分類結果は図 4 のように得られる。各画像のスコアが最も高いクラスを画像のクラスに採用している。結果をみると、学習時の正確性はほぼ 100%，テスト時の正確性は約 75%であることが分かる。街並みの特徴を抽出した画像の分類が一定の精度で可能であることが分かった。人でも時間を要すれば街並みのクラス分類を同程度の精度で実行することは可能であるが、高速に学習し、人に依存せずに推定を再現できるという観点をふまえると、比較的高い精度と考えられる。

3. 2 街並みに対する訪問意欲の分類 (MODEL_01)

建築に関する画像の分類が上述の精度で可能であることを確認したため、次に感性評価による街並みの分類を行った。データセットは、表 1 から 10 箇所を選定^{注6)}、した後に無作為に選定した 10 枚、計 100 枚の画像である。テスト用の画像は先述した各都市 100 枚の中から無作為に 5 枚(学習用とテスト用の重複はなし)、計 50 枚とした。分類クラスは「訪問したい」「訪問したくない」という感性評価である。感性評価は一名の被験者(建築系学科所属の 4 回生)による被験者実験の結果を用いている。被験者実験では直感で画像をクラスに分類して貰った。通常このような被験者実験には複数人の代表値を用いることが多いが、個人の個人的な感性評価にどれだけ近づけるかということに重きを置いているため被験者は一人とした。

結果を図 5 に示す^{注7)}。これをみると「訪問したい」の確率が 50%から 55%に収束してしまう結果となった。どちらでも無いという結果も考えられるが、学習させるデータセットの画像枚数を増やし、再度、感性評価による街並みの分類を行った。

3. 3 街並みに対する訪問意欲の分類 (MODEL_02)

先述と同様に再度感性評価による街並みの分類を行った。今回のデータセットは、3.1 で使用したデータセットから、3.2 で使用したテスト用のデータを抜いたものに、各都市の新たなストリートビュー画像をそれぞれに補充した計 2100 枚である。テスト用のデータセットは 3.2 で使用したテスト用データセットと同じものである。分類クラスは 3.2 と同じ「訪問したい」「訪問したくない」の 2 種類であり、3.2 と同様に被験者実験を行った。学習結果を図 6 に示す。今回は学習時とテスト時、ともに 90%以上の正確性を持つ結果となった。次章においてこの被験者の感性評価の推定結果について統計的仮説検定を用いて検証する。

4. DeepLearning による訪問意欲の推定の検証

MODEL_01・MODEL_02・被験者の感性評価(図中では「Human」と表記)を比較する。テスト用の画像に対して、DeepLearning を用いた推定においては「訪問したい」と分類するスコアが 0 から 100%の連続変数として示される。これに対応して、テスト画像に対しては同じ被験者に「訪問した」度合いを 10 段階で評価して貰った。この値を本章では比較した。図 7 に三者の箱ひげ図を示す。まず MODEL_01 において、中央値(約 50)が HUMAN の中央値(約 70)と差が大きい結果となっ

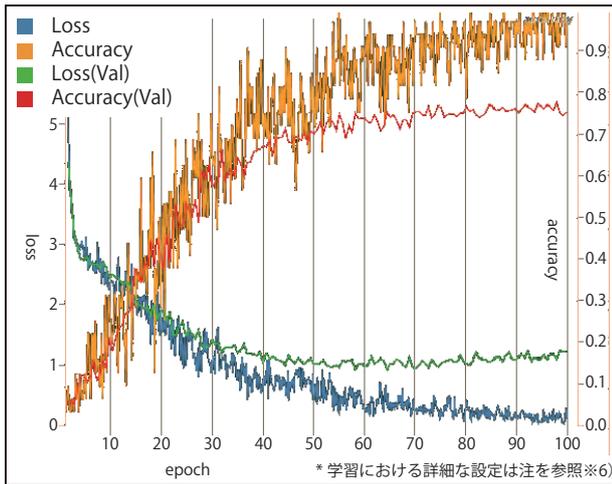


図3 学習結果_都市の街並みの分類

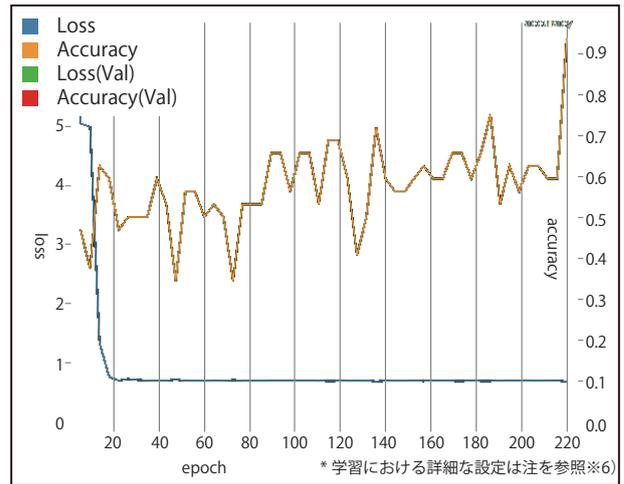


図5 学習結果_感性評価による分類 (MODEL_01)

	rank	name	score
	1	Canada	73.3
	2	Portugal	26.1
	3	South Africa	0.5
	4	Czech Republic	0.1
	rank	name	score
	1	America	96.9
	2	Italy	3.0
	3	Hong Kong	0.1
	4	Japan_Osaka	0.0
	5	Czech Republic	0.0

図4 画像の分類結果の一例

た。また値の散らばりがほとんど見られないことが分かる。狭い範囲においてHUMANの値の分布と類似の傾向を有している可能性はあるが、図8のヒストグラムを見ると視覚的に分布の形状が明らかに異なることが分かる。MODEL_01は一般性を統計的に検証するまでもなくHUMANの感性評価を推定出来ていないと考えられる。

次にMODEL_02を見ると、中央値がほぼHUMANの中央値と同値(約50)であることがわかる。このことからMODEL_02は、標本の結果としては標本間の誤差はあるが高い精度で全体の中央値を推定できていると言える。この結果の一般性については統計的仮説検定を用いて検証する。次に、MODEL_02は四分位の範囲の重なりも広いこともわかる。また両者のヒストグラムを見ても形状が似ていることが視覚的に読み取れる。このことからMODEL_02は、標本の結果としては標本間の誤差のあるがHUMANの印象評価を全体の傾向としての推定出来ている可能性がある。この結果についても統計的仮説検定により検証する。なお今回、視覚的な分析においてMODEL_01の推定精度は低いことが明らかのため以降は省略する。

統計的仮説検定による結果を後述する。MODEL_02とHUMANの差について検証するために効果量を算出した。図9より、ヒストグラムから正規性は予測されないと判断できることからノンパラメトリックとして両者の

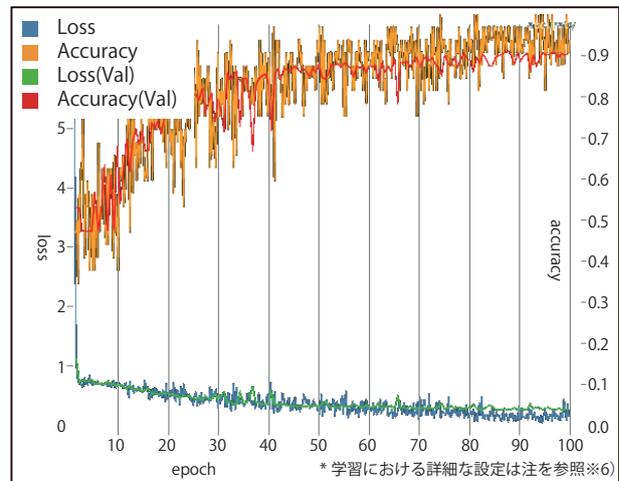


図6 学習結果_感性評価による分類 (MODEL_02)

差の効果量(対応あり)を算出した結果、0.12(Cliffのdだと0.0152)となった。心理学的な研究において0.2以下を小さいと言及することを勧察すると両者の差は小さいということを肯定できると考えられる。次に統計的仮説検定であるが、標本結果の母集団・一般性に適用の解釈において有意差が認めれたとしても、差は微小であることからWilcoxonの符号付順位検定の実施は省略した。

次に値の大小の傾向の推定について検証する。この検証においてもノンパラメトリックと考え、スピアマン(spearman)の順位相関係数を算出した結果、0.4699となった(図9)。この値は心理学的な研究においては中程度の相関関係と言及される。しかし、標本結果の母集団・一般性に適用の解釈において無相関である可能性があるため、無相関仮説の検定を行なった。p値は0.0005742となり1%水準においても高度に有意な結果となった。このことから相関関係の母集団の適用可能性・一般性を肯定できると考えられる。

5. 考察とまとめ

DeepLearningを用いて作成した感性評価推定MODEL_02の推定精度について以下のように考察する。

Wilcoxon の順位和検定は実施しなかったが、MODEL_02 と HUMAN の効果量 0.12 となったことから、標本結果の母集団・一般性に適用の解釈において有意差が認めれたとしても、MODEL_02 は HUMAN の中央値を高い精度で推定できていると考えられる。値の大小の傾向の推定については、スピアマンの順位相関係数が 0.4699、無相関仮説の p 値が 0.0005742 となったことから、母集団・一般性に適用の解釈においても順位としての相関関係を有すると考えられ、中程度に推定できていると考えられる。

以上のように本稿では、研究で用いた街路を対象にはあるが上記のような精度を有する街並みに対する感性評価の推定モデルを構築することができた。また同時に DeepLearning という建築・都市デザイン分野においては未検討な技術の適用可能性の一端も示すことができた。今後は都市を変えた検証，人を変えた検証，学習モデル・枚数・学習層を変えた検証，街並み以外の対象について取り組む予定である。

注

- ※ 1) 下記 URL はストリートビューの候補に選定するために参考にした文献。
<https://wondertrip.jp/zekkei/9005.html>
<http://japontimes.livedoor.biz/archives/27655746.html>
<https://matome.naver.jp/odai/2133631453308404201?page=1>
- ※ 2) 中国、インド、中東地区、アフリカ、ドイツなど Google ストリートビューで閲覧不可能な場所は候補から除外。
- ※ 3) 学習に使う画像データが一定量 (2100 枚) 以上ある場合、データを学習用とテスト (Validation) 用に分けて実行が可能のため 2100 枚を目安として。また街並みの画像であるため、画像が単一の建築物・地面・空に占められることのないように配慮しながら作成した。
- ※ 4) MachineLearning を用いた画像認識において、特徴抽出処理にニューラル・ネットワークを用いることにより、高精度な認識が可能である。近年では、ニューラル・ネットワークの中でも畳み込みニューラル・ネットワークを用いて、画像認識を行う手法が注目されている。畳み込みニューラル・ネットワークでは、各ニューロンが入力の全画素に接続されるのではなく、画像の一部の近傍領域のみに接続される。ここで、結合重みが画像の領域ごとに変化しない場合、この接続は画像の一部にフィルタを畳み込むのと同義である。このような構造が、全接続のネットワークよりも位置ずれに強い一因とされる。畳み込みニューラル・ネットワークでは、局所受容野によって入力画像の一部、または一部の特徴の出力にしか接続されていない複数のネットワークを、最終段において全接続のネットワークを用いて結合する。畳み込み層では、最後に活性化関数を用いて出力が得られる。従来の畳み込みニューラル・ネットワークにおいては、局所受容野層において一般に複数のフィルタを用いている。一方、Network In Network では、単純なフィルタの代わりに小規模な多層パーセプトロンを組み込む。これにより、従来の畳み込みニューラル・ネットワークよりも複雑な表現が可能になり、画像の特徴抽出能力の向上が可能になる。
- ※ 5) 図 3, 5, 6 の学習条件はそれぞれ、Frame work:chainer, Batch Size:32, Train Type:Image Classification, Network:nin.py, Resize mode:squash, Color mode:RGB, Epoch はそれぞれ図 3 は 100, 図 5 は 220, 図 6 は 100 である。
- ※ 6) 選定した場所は、アメリカ、カナダ、チェコ、イギリス、香港、イタリア、日本 (京都)、スコットランド、南アフリカ、スペイン
- ※ 7) データセットの画像枚数が少ないため、全てを学習に使用している。そのため学習時の正確性のみ表示している。

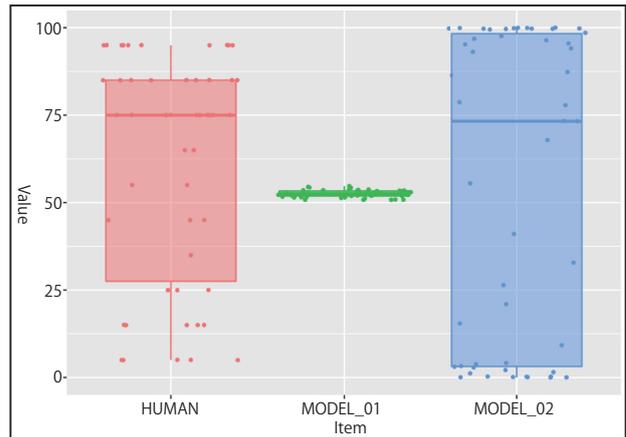


図 7 感性評価実験の結果と推定値の箱ひげ図

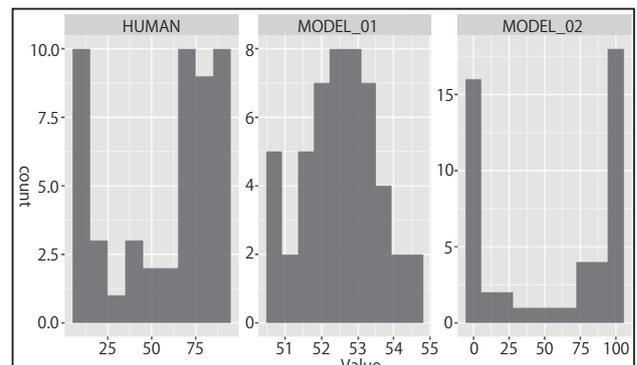


図 8 感性評価実験の結果と推定値のヒストグラム

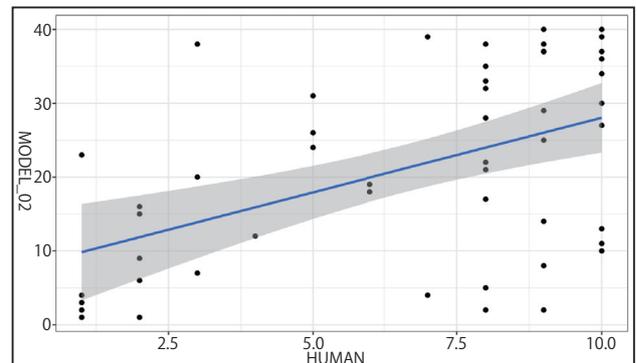


図 9 HUMAN と MODEL_02 の順位の散布図

参考文献

- 01) 山田悟史, 藤井健史, 宗本晋作: 全方位緑視率を用いた緑地環境に対する印象評価推定モデル作成と検証 - 全方位緑視率と印象評価実験をもとにした自己組織化マップを用いて -, 日本建築学会計画系論文集, 第 81 巻, 第 727 号, pp. 2083-2093, 2016. 9, 日本建築学会
- 02) 宗本晋作・山田悟史: ベイジアンネットワークを用いた緑視率に基づく緑環境評価の確率モデルに関する研究, 日本建築学会計画系論文集, 第 81 巻, 第 729 号, pp. 2411-2419, 2016. 11, 日本建築学会
- 03) 古井秀弥, 黒木修隆, 廣瀬哲也, 沼昌宏: 「事前教師あり学習を適用した Network in Network による画像認識の高精度化」第 14 回情報科学技術フォーラム, 第 3 分冊, pp. 179-pp. 180, 2015
- 04) 大久保街亜, 岡田謙介: 伝えるための心理統計: 効果量・信頼区間・検定力, 2012. 1, 勁草書房
- 05) 清水亮: はじめての DeepLearning プログラミング, 技術評論社
- 06) 藤田一弥, 高原歩, 株式会社フォワードネットワーク: 実装 DeepLearning, 2016. 11, Ohmsha

* 1 立命館大学 理工学部 建築都市デザイン学科

* 2 立命館大学 理工学部 建築都市デザイン学科 講師・博士 (工学)