

ニューラルネットワークを用いた建設プロジェクト初期段階における概算工事費の予測方法に関する検討

A Study on the Way of Prediction for Rough Construction Cost in the Early Stage of the Project Using Neural Network Techniques

○加藤 泰正*¹, 伊藤 一義*², 村瀬 弘幸*³
Yasumasa Kato*¹, Kazuyoshi Ito*² and Hiroyuki Murase*³

- *1 株式会社日本設計 PM・CM部 シニアエンジニア 博士(工学)
NIHON SEKKEI, Inc. Senior Project Designer, Project Consulting Management Dr. Eng.
- *2 株式会社日本設計 PM・CM部 副部長
NIHON SEKKEI, Inc. Senior Director, Project Consulting Management
- *3 株式会社日本設計 コスト設計部 主管 修士(工学)
NIHON SEKKEI, Inc. Senior Project Cost Planner M. Eng.

キーワード: ニューラルネットワーク; 過学習; コストマネジメント; 概算工事費; 予測システム
Keywords: Neural network; overfitting; cost management; rough construction cost; prediction system.

1. はじめに

建設プロジェクトが巨大化、複雑化する中、情報量が極めて少ない初期段階においても、事業性を評価する観点から目標予算(概算工事費)を定めてコストマネジメントに着手する。一般的には、過去の類似案件のコストを分析、補正して対応するが、諸条件が異なるため、必ずしも適切な目標予算を設定できるとは限らない。一方、組織内において工事費実績データが蓄積されてきており、それらを活用した汎用性の高い概算工事費予測システムの開発が望まれる。このような予測システムは、不確実性が高い建設プロジェクト初期段階において、セカンドオピニオンという位置付けでの利用が期待される。

蓄積された実績データから概算工事費を予測する試みは古くからあるが^{1),2)}、分析上同質のデータを扱っているため、施設用途を特定する他、立地場所及び着工年に基づく物価変動による影響は限定することになる。このため、汎用性の高い予測システムへの展開は難しい。

本報は延床面積当りの概算工事費(以下、「概算単価」と呼ぶ)予測システム開発の前段階として、ニューラルネットワークを用いた予測手法³⁾に対し、利便性と予測精度を高めるための方法を幾つか検討したので報告する。

2. 方法

2.1. 検討概要

検討事項は3点とし、以下に各概要について説明する。
検討1: 利便性を高めるため、単一のニューラルネットワーク予測器において、複数の施設用途及び複合施設を対象にすることを考える。そのためには、入力する施設用途と

いうカテゴリーデータをワンホット表現³⁾(対象は1、そうでない場合は0)して扱う。複合施設の場合はそれらの入力データの対象用途の合計が1になるように構成比率を設定する。一方、予測範囲を限定するため、建物の仕様に基づくグレード^{4),5)}を数値化して入力する。本報で用いるグレードは工事単価の統計値を利用して定義する。

検討2: 検討1に示すように予測器の適用範囲を拡大する場合、予測精度を高めるために、学習データを幅広く収集しなければならない。しかし、工事費は立地場所や着工時の需給バランスによって変動するので、それらの影響を除去する必要がある。そこで、立地場所と着工年を入力データとして加え、それらを起因とする物価変動の状況を学習させることを考える。具体的には、立地場所は、県・政令指定都市別の平均地価⁶⁾及び県別の工事費指数⁷⁾を用いる。着工年は西暦を直接数値データとして入力する。これはニューラルネットワークが非線形性を考慮できることに期待している。さらに、全国平均的な物価変動のトレンドを学習させるため、建設工事費デフレータ(建築総合指数⁸⁾)の年度別勾配を入力する。

検討3: ニューラルネットワークは学習能力が高い一方、過学習によって予測精度を落とすことがある。今回は適用範囲の拡大を考える中で、学習データが偏在する恐れがあり、過学習の対策は特に重要となる。ここでは予測値と実績値の比較による検証に加え、予測値のみからなる妥当性確認を行う。この妥当性確認は入力データをパラメトリックに変化させ、予測値に非現実的な変化が生じないかを観察する方法とする。検証のような明確さはないが、過学習を判断する重要な手段の一つになると考える。

2.2. データ概要

収集したデータは立地場所が国内各地域、着工が2000年から2019年の計233件であり、種々の施設用途、グレード、規模、構造種別を含んでいる。施設用途とグレードの関係下における各データ件数を表1に示す。

同表内の施設用途01から09は再開発でよく見られる用途であり、02から06は複合施設である。グレードは全用途における延床面積当りの工事費（消費税別の見積額）に関する統計値を利用して設定した（表1下段参照）。用途別にデータ件数を見ると、1件のものを除けば、グレードは概ね1から4段階の分布になっており、人間が経験に基づき直感的に設定できる範囲である。

2.3. 検討手順

ニューラルネットワークの入出力データ項目の詳細を表2に示す。カテゴリーデータを除く数値データは、入力には標準化（平均0、標準偏差1）し、出力（教師データ）は最大値が1になるように基準化して用いる。

本計算はプログラミング言語 Python を用い、深層学習ライブラリ Keras⁹⁾を利用した。ニューラルネットワークの中間層は3層とし、各中間層の活性化関数は双曲線関数 (tanh)、出力層はシグモイド関数を用いた。損失関数は二乗和誤差を使用し、学習の最適化アルゴリズムは adam とした。また、過学習を抑えるため、損失関数に L2 正則化項を考慮した³⁾。

「検討1」は、次の3ケースのデータ項目対象（表2右列の○印参照）において学習を行う。

Case1：再開発系事務所（用途01）

Case2：再開発系4用途（用途01から04）

Case3：全件（用途01から12）

表3には出力データに関する統計値を示している。対象用途及びデータ数が増加すると、変動係数も増加するという特徴がある。なお、表1に示すデータ件数は、再開発系以外のデータを含んでいるため、Case1,2とは一致しない。

検証方法は、各ケース、全データの15%程度を検証データとして無作為抽出し、その内80%以上が次式を満たすかを判定する。

$$-0.15 < \frac{y' - y}{y} < 0.15 \quad (1)$$

ただし、y':概算単価予測値、y:概算単価実績値とする。

以上の準備の下、Case1,2,3において、予測値における平均二乗和誤差の比較を行う。さらに、事務所の検証データのみは各ケースで共通にしておき、その平均二乗和誤差を比較することで、他用途データの学習への貢献度を考察する。参考に、各ケース同データを用いた線形重回帰分析の結果とも比較する。

「検討2」はCase2の学習結果を用いて、立地場所と着工年の違いによる予測値の変化を示す。これにより、立地場

表1. 施設用途とグレードの関係下における各データ件数

施設用途	グレード						合計	構成比
	I	II	III	IV	V	V強		
01事務所		13	19	4	1		37	15.9%
02事務所・商業		4	6	6			16	6.9%
03事務所・住宅		1	5	1			7	3.0%
04事務所・商業・住宅		2	2	1	1		6	2.6%
05事務所・宿泊			3	1			4	1.7%
06商業・住宅	1	5	2				8	3.4%
07住宅		9	5	2	1		17	7.3%
08宿泊		3	2	2			7	3.0%
09商業	2	6	1				9	3.9%
10庁舎		1	12	2			15	6.4%
11文化		1	3	3	2	5	14	6.0%
12スポーツ	1	1	2	3			7	3.0%
13教育・研究		20	27	5	5	1	58	24.9%
14医療		3	8	2			13	5.6%
15福祉				3			3	1.3%
16生産	1		2	3	1	2	9	3.9%
17倉庫・物流	2	1					3	1.3%
合計	7	70	99	38	11	8	233	100.0%
構成比	3.0%	30.0%	42.5%	16.3%	4.7%	3.4%	100%	—

グレード I : A-1.5σ未満、II : A-1.5σ以上 A-0.5σ未満、
 III : A-0.5σ以上 A+0.5σ未満、IV : A+0.5σ以上 A+1.5σ未満、
 V : A+1.5σ以上 A+2.5σ未満、V強 : A+2.5σ以上
 A : 延床面積当りの工事費の平均値（工事費は2018年、東京での補正金額を使用。補正には国土交通省の建設工事費デフレータ及び地域別工事費指数を用いた）
 σ : 同 標準偏差値

表2. ニューラルネットワークの入出力データ項目と計算ケース (Case1,2,3)

分類	データ項目名	データ種別	備考	Case 1	Case 2	Case 3
入力	延床面積	数値	・最小値:166㎡、最大値:24万㎡	○	○	○
	地上階数	数値	・最小値:1階、最大値:52階	○	○	○
	地下階数	数値	・最小値:0階、最大値:5階	○	○	○
	用途01-事務所	0 or 1		—	○	○
	用途02-商業	0 or 1		—	○	○
	用途03-住宅	0 or 1		—	○	○
	用途04-宿泊	0 or 1	・用途種別はカテゴリーデータであり、対象を1、	—	○	○
	用途05-庁舎	0 or 1	そうでない場合を0と数値化して用いる。	—	—	○
	用途06-文化	0 or 1	・複合施設の場合は、最大3用途までとし、対象用途の合計値が1になるように設定する。	—	—	○
	用途07-スポーツ	0 or 1	・Case1は再開発系事務所を対象とする。	—	—	○
	用途08-教育・研究	0 or 1	・Case2は再開発系4用途(事務所、商業、住宅、宿泊)を対象とする。	—	—	○
	用途09-医療	0 or 1		—	—	○
用途10-福祉	0 or 1		—	—	○	
用途11-生産	0 or 1		—	—	○	
用途12-倉庫・物流	0 or 1		—	—	○	
出力	グレード	数値	・全用途における2018年東京の換算工事費の延床面積当りの工事費を次の6段階に設定する。 ・グレードは以下のように数値化する。 I : 1、II : 2、III : 3、IV : 4、V : 5、V強 : 6	○	○	○
	構造1-RC	0 or 1	・構造種別はカテゴリーデータであり、対象を1、	○	○	○
	構造2-S	0 or 1	そうでない場合を0と数値化して用いる。	○	○	○
	構造3-SRC	0 or 1	・複合構造は考慮せず、代表の構造種別を設定する。	○	○	○
出力	2018年平均地価	数値	・国土交通省 土地・建設産業局の土地総合情報システムより、2018年における都道府県別（政令指定都市を除く）及び、政令指定都市別の商業地の土地面積当りの平均取引額を使用する。東京都は23区内と外に分ける。 ・最小値:17千円/㎡、最大値:1,375千円/㎡	○	○	○
	2018年地域別工事費指数	数値	・国土交通省、平成30年度 新営予算単価より、県別の教育鉄筋コンクリート造の地域別工事費指数を用いる。 ・最小値:95、最大値:104	○	○	○
	着工年	数値	・着工年の西暦を直接数値データとする。 ・最小値:2000年、最大値:2019年	○	○	○
	デフレータ過去2年勾配	数値	・国土交通省、建設工事費デフレータより、建築総合指数を参照し、現在年の指数を2年前の指数で除した値を用いる。 ・最小値:0.97、最大値:1.06	○	○	○
出力	延床面積当りの工事費(概算単価)	数値	・場所と時間による物価補正は行わない。 ・最小値:57千円/㎡、最大値:887千円/㎡	○	○	○

○:対象

所及び着工年の影響を学習できているかを考察する。

「検討3」は Case1 の学習結果を用いて、過学習の例を示す。さらに、妥当性確認の例として、Case2 の過学習を抑えた結果を用い、延床面積の違いによる予測値の変化を提示する。

3. 考察

3.1. 「検討1」について

Case1,2,3 における全データ (学習データ及び検証データ) の予測値における平均二乗和誤差を図1に示す。Case3 の線形重回帰分析のみは、2.3.に示した検証条件を満足できなかったが、その他は全て満足している。

同図に示す誤差は、概ね変動係数に沿って大きくなっているが、Case1 と Case2 の違いは極めて小さい。これは、複数用途の組み合わせにもよるが、複数用途の予測ができることを示唆している。

次に、再開発系事務所の検証データのみにおける平均二乗和誤差の関係を図2に示す。Case2,3 の誤差が Case1 よりも小さくなっているため、事務所においては、他用途の学習がプラスに働いたと考えることができる。ただし、Case3 は全データの誤差が大きいため、事務所以外の用途では反対の結果も起こり得る。

また、図1,2には、線形重回帰分析の結果も合わせて示している。図2のCase2を除き、ニューラルネットワークの方が、誤差が小さいという結果を得た。

さらに、Case2 の学習結果を用い、2用途から成る複合施設の延床面積構成比の違いによる概算単価予測値の変化を図3に示す。2019年A市で着工、延床面積40,000㎡、地上20階、地下2階、鉄骨鉄筋コンクリート造、グレードⅢの施設を想定している。ここでは、用途を事務所/商業、事務所/住宅と、2用途間で構成比を0.25刻みで変化させ、概算単価予測値の変化状況を示している。何れも2用途間で線形補間的な予測値が得られるという結果を得た。

以上より、本予測において、複数用途及び複合施設を扱うことが可能であることを示した。

3.2. 「検討2」について

Case2 の学習結果を用い、立地場所別概算単価予測値の比較を図4に示す。何れも2019年着工、延床面積40,000㎡、地上20階、地下2階、鉄骨造、グレードⅢの事務所を想定している。地価は高い順に、A市、B市、C市、D市、E市、F市となっており、この順に、地域別工事費指数は100, 101, 96, 99, 95, 100となっている。

予測値は、地域別工事費指数未考慮の場合、地価の影響をほとんど受けていない。考慮の場合は、当指数の大小関係が反映されており、当指数の効果が学習されていると考えられる。本計算では、地価よりも地域別工事費指数の影響が大きいという結果を得た。

表3. 計算ケース (Case1,2,3) における延床面積当りの工事費データに関する統計値一覧

項目	Case1	Case2	Case3
データ件数	36	100	233
平均値(円/㎡)	338,397	336,885	367,310
標準偏差(円/㎡)	90,901	105,403	141,459
変動係数	0.269	0.313	0.385

注意事項:
1) 立地場所及び着工年による工事費の補正はしていない。
2) 変動係数=標準偏差/平均値

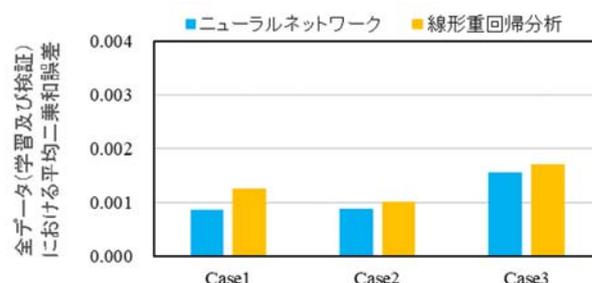


図1. 全データ (学習及び検証) における平均二乗和誤差

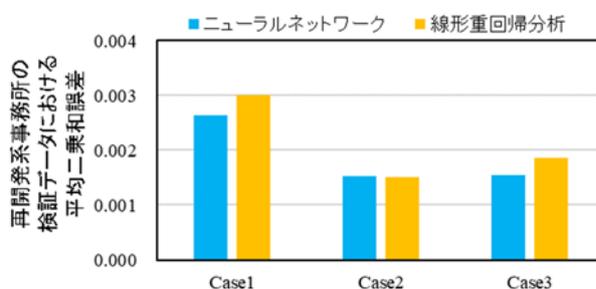


図2. 再開発系事務所の検証データにおける平均二乗和誤差

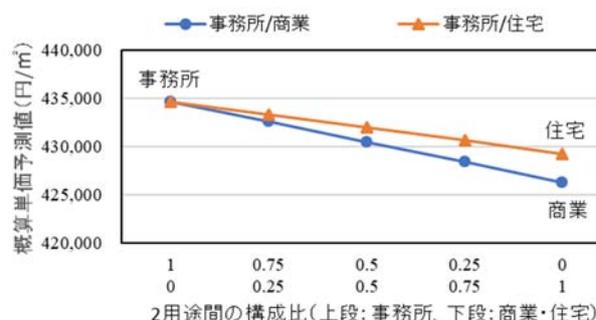


図3. 2用途から成る複合施設の延床面積構成比別概算単価予測値の変化

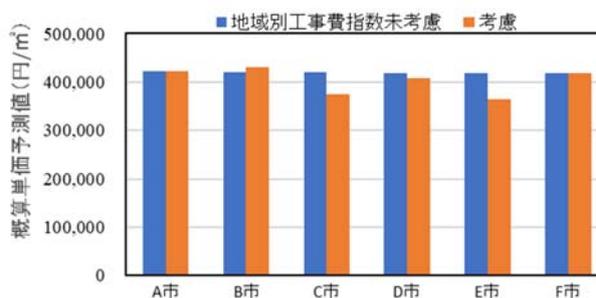


図4. 立地場所別概算単価予測値の比較

次に、概算単価予測値の着工年別推移を建設工事費デフレータ⁸⁾と共に図5に示す。前モデルで立地場所はA市としている。予測値はデフレータよりも後半やや高めであり、本ケースは社会的要求水準の上昇があったと示唆される。

以上より、ニューラルネットワークは、立地場所及び着工年による物価変動の状況を学習できることを確認した。

3.3. 「検討3」について

Case1の学習結果を用い、2019年A市で着工、延床面積200,000㎡、地上40階、地下2階、鉄骨造の事務所を想定し、グレード別概算単価予測値の変化を図6に示す。過学習の影響を見るため、損失関数にL2正則化項未考慮の場合と考慮の場合とを比較している。なお、未考慮の場合においても、2.3.に示す検証条件を満足させている。

本事例はグレードの高い領域で学習データ量が少なくなっており(表1参照)、その領域では特定のデータの特徴しか学習できていない可能性がある。L2正則化項未考慮の場合、考慮の場合よりもグレードの影響が顕著に小さくなっていることから、当結果は過学習になっていると推測される。

次にCase2の学習結果を用い、グレード別及び延床面積別概算単価予測値の変化を図7に示す。2019年A市で着工、地上40階、地下2階、鉄骨造の事務所を想定している。本計算例からは、グレードが上がる程、延床面積の増加における概算単価の減少率が、わずかであるが小さくなるという結果を得た。

このようなパラメトリックスタディを種々行うことは、コスト特性の理解を深めると共に、過学習の評価の一助となり、ニューラルネットワークによる予測値の妥当性確認ができると考えられる。

4. おわりに

ニューラルネットワークを用い、建設プロジェクト初期段階における概算単価の予測方法を、利便性と予測精度を高めるという観点から幾つか検討した。得られた知見は以下の通りである。

- 1) 単一の予測器の中で、異なる施設用途を対象にして、一定の検証条件を満たす概算単価を予測することができた。
- 2) 立地場所及び着工年の違いによる物価変動の状況を学習させることができた。
- 3) 予測値と実績値を比較する検証に加え、予測値だけのパラメトリックスタディがコスト特性及び過学習の評価の一助となり、予測値の妥当性確認ができた。

これらはニューラルネットワークの有用性を示すものである。今後は一部施設用途の細分化、学習データ量の拡大、入力項目の見直し等、さらに検討を進めることで、汎用性の高い予測システムに展開したいと考えている。

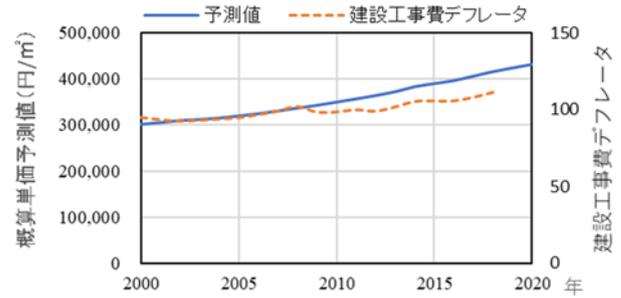


図5. 概算単価予測値の着工年別推移

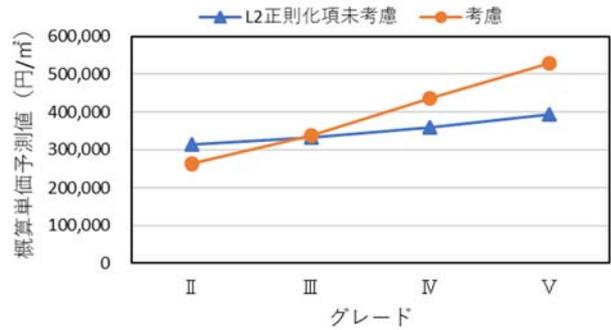


図6. グレード別概算単価予測値の変化

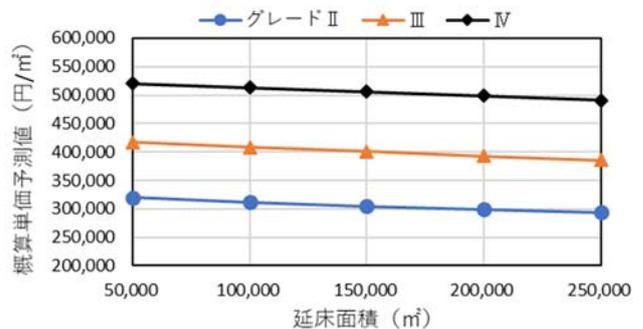


図7. グレード別及び延床面積別における概算単価予測値の変化(妥当性確認の一例)

謝辞 本研究を進めるに当たり、日本設計 小坂幹氏、柳泰彦氏には貴重な意見を頂きました。深く感謝致します。

[参考文献]

- 1) 石原徹也:多変量解析による企画段階での建築コストの予測に関する研究, 日本建築学会関東支部研究報告書, 421-424, 1986
- 2) 具源龍, 加藤直樹:データマイニング手法を用いた建築プロジェクトの工事費予測モデルの提案, 日本建築学会環境系論文集 第598号, 79-85, 2005.12
- 3) Ian Goodfellow 他:「深層学習」, KADOKAWA, 2018.2
- 4) 一般財団法人建築物価調査会総合研究所: <https://www.jbci.jp>
- 5) 田中修一:「デザイナーのための建築事業収支計画」, 彰国社, 2007
- 6) 国土交通省土地・建設産業局: 土地総合情報システム
- 7) 国土交通省大臣官房長官官舎部: 平成30年度新当予算単価
- 8) 国土交通省総合政策局情報政策課: 建設工事費デフレータ
- 9) <https://keras.io/>