千葉商大論叢　第 58 巻　第 2 号（2020 年 11 月）pp. 81-93

〔論　説〕

機械学習を利用した注意機構を持つ回帰モデルによる影響度分析

内　海　幸　久

1　序

回帰分析は，定量的なデータが豊富な経済学，経営学，政治学など多くの社会科学の分野で必要不可欠な分析手法となっている．近年では，数量化 I 類の手法によって，定性的なデータを含む状態でも回帰分析が可能となり，多くの分野で回帰分析が利用される．その回帰式の解釈も多岐にわたる．アンケート調査やテキストマイニングにおいては，係数を被説明変数の影響力として捉えることが多い．実際，線形の重回帰モデルは，l 変数を持つアフィン関数 $y = \sum_{i=1}^{l} a_i x_i + b$ を利用して被説明変数を近似するものである．このため，アンケートデータのように入力データが同一の変域に収まるデータであるならば，係数の a_i は，第 i 変数の被説明変数への影響力であると解釈される．このことから，入力データの変数の変域がそれぞれ異なる状況でも，変数を正規化することで，回帰式の係数に影響力の解釈を与えることができる．

その特徴は，二種類にわけられる．回帰係数は，負値を取ることもあり得る．しかし，影響力として，係数の絶対値を利用することで，その問題を回避できる．つまり，正値方向も，負値方向も絶対値という均一の重み付けを利用することで，影響力の順位を知ることが可能となる．これが，第一の特徴である．第二は，モデルとしての大きな特徴を係数という形で，捉えている点である．つまり，モデル全体としての変数の影響力を分析できることが，第二の特徴である．

回帰係数は，その絶対値の大小によって被説明変数への影響力がわかるものの，確率分布のような割合を表す精緻な影響度ではないこと，また，データ毎の影響力が未知であることなど，幾つかの問題も抱えている．

本稿の目的は，この二点の問題点を解決するべく，自然言語処理で近年利用されている，注意機構を導入した回帰分析を提案することである．注意機構を導入することによって，回帰係数の影響力を割合として表しつつ，かつ，個別のデータの変数の影響度も同時に知ることが可能となる．注意機構は，もともと，自然言語の分野で Vasawani，et al (2017) らによって，提案がなされたネットワーク構造である．その後，多くの自然言語処理の手法に応用された．注意機構を用いたネットワーク構造は，その大半が，自然言語処理の分野で応用されているが，Devlin，el al (2018) の BERT 以降，分類問題でも，応用されるようになった．画像認識の分野においては，Fukuni，et al (2019) らによって注意機構が利用されている．

本稿では，注意機構をネットワーク構造に導入した回帰モデルを構築する．主要な帰結は，以下の三点にまとめられる．第一は，注意機構付き回帰分析を行うことで，全体としては影響力があるように見えるが，個別データではそれほど影響力がない個別データの存在が確認されたことである．第二は，同様に，全体としては影響力は小さいが，個別データにおいては，大きな影響力を持っている個別データの存在も確認されたことである．つまり，モデル全体での影響力と個別データごとには，変数間の影響力の違いが見られる．個別データでの影響力を求められることが，機械学習を利用した重回帰モデルの特徴と言え

る．具体的には，個別データ毎に，影響力のある変数を求めることができるため，そのデータに応じた予測や対応なども可能となる．第三は，影響度が低くなることで自動的に多重共線性が回避される点である．

2 章において本稿で展開する注意機構付き回帰モデルを定義や解釈を述べる．3 章において scikitlearn から入手できるボストン住宅価格データを利用して，注意機構付き回帰モデルの特徴を紹介しつつ，4 章にて帰結を述べる．

2　注意機構を持つ重回帰モデル

2.1　ネットワーク構造

バッチサイズを n とし，その集合を $N = \{1, ..., n\}$ で表す．入力データは l 種類の特徴を持つ l 次元ベクトルで表記されるとし，$k \in N$ 番目のデータは，$(x_1^k, ..., x_l^k) \in \mathbb{R}^l$ によって記述される．変数毎に，データの桁が異なる可能性があるために，入力データについて，正規化を施す．バッチ正規化を関数 b とおくと，バッチ正規化された値は，$b(x) = (b_1(x_1), ..., b_l(x_l))$ と記述される．

本稿で提案する注意機構付き重回帰モデルは，この正規化されたデータに確率ベクトルをかぶせることで，データのどの項目が被説明変数に対して，重要な要因になるのかを明らかにする．注意機構の構成について述べる．注意ベクトルとは，所与の l 次元ベクトル $\alpha = (\alpha_1, ..., \alpha_l) \in \mathbb{R}^l$ と正規化されたデータ値の要素積にソフトマックスを施した，確率ベクトルのことである．具体的には，任意の $j = 1, ..., l$ に対して，

$$m_j = \frac{e^{\alpha_j b_j(x_j)}}{\sum_{k=1}^{l} e^{\alpha_k b_k(x_k)}}$$

と定義されるものである．$m = (m_1, ..., m_l)$ と表記する．所与の $\alpha = (\alpha_1, ..., \alpha_l)$ は誤差を最小にするように求められ，これに基づき，注意ベクトルが計算されることとなる．

注意機構付き重回帰モデルは，この注意ベクトルと正規化されたデータの要素積を入力データとした回帰モデルといえる．全結合 1 層タイプは $m \otimes b(x) = (m_1 b_1(x_1), ..., m_n b_l(x_l))$ を全結合層の入力データ，全結合層の重みを $(w_1, ..., w_l) \in \mathbb{R}^l$，バイアスを $c \in \mathbb{R}$ とおくと，

$$y = \sum_{j=1}^{l} w_j m_j b_j(x_j) + c$$

と求まる．確率 m_j で変数 j に注意を向けるという意味から，m_j は変数 j の影響度を表していると解釈できる．これらをまとめると，全結合層 1 層の式は，

$$y = \sum_{j=1}^{l} w_j \frac{e^{\alpha_j b_j(x_j)}}{\sum_{k=1}^{l} e^{\alpha_k b_k(x_k)}} b_j(x_j) + c$$

と求まる．

回帰式としては，非線形の構造になるが，本稿では，便宜上，注意機構による重みの抽出後の全結合層が 1 層のモデルを線形モデル，2 層以上のモデルを非線形モデルと呼ぶこととする．

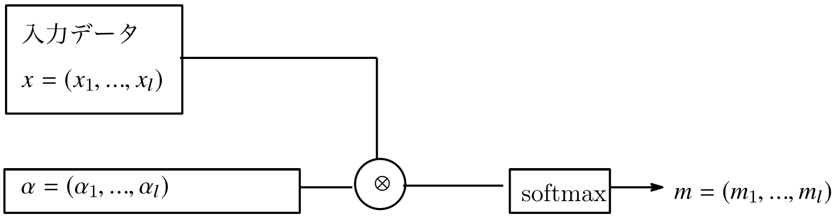

図 1: 注意ベクトル

2.2　注意機構の解釈

注意機構の利点や特徴は，3 点にまとめられる．第一は，注意ベクトル $(m_1, ..., m_l)$ によって，データごとの被説明変数への影響度がはっきりとわかることである．第二は，重回帰分析によく観察される多重共線性問題を自動的に解決する点である．第三は，注意ベクトルによる影響度・注目度という解釈可能な非線形の重回帰モデルを平易に構築できる点である．

注意ベクトル $(m_1, ..., m_l)$ は，確率ベクトルであるので，各教師データに対して，l 種類ある入力データの j 番目の変数に対して，m_j の確率だけ影響力を持つと解釈できる．加えて，データ毎に，ソフトマックス関数によって，

$$m_j = \frac{e^{\alpha_j b_j(x_j)}}{\sum_{k=1}^{l} e^{\alpha_k b_k(x_k)}}$$

と計算される．これより，注意の確率は，データ毎に教師データに対してどれくらいの影響力があるのかの，個別の指標を与えることになる．例えば，注意の値が 0 であるならば，その項目は，教師データにとって重要でない変数になることがわかる．注意の値が，相対的に大きい項目は，教師データにとって重要な変数となることが示唆される．注意ベクトルの確率は，回帰係数の解釈としてではなく，影響度としての本来的な意味を与えることとなる．

全体の影響度ではなく，個別のデータ毎の影響度が求まることの重要性は，無相関だが規則性があるデータに注目できる点である．表 1 のデータを考えてみる．図 2 のグラフのように項目 1 で，1 番を選択した人は，項目 3 で，3 番を選択している人が多い．一方，項目 1 で，1 以外を選択している人の項目 3 での選択は，様々である．この場合，全体としては無相関に近いが，明らかに，項目 1 の 1 番と項目 3 の 3 番には，規則性が見られる．

実際，項目 1 と項目 3 の相関係数を計算すると，ほぼ 0 となり，無相関となる．しかし，項目 1 の 1 番と項目 3 の 3 番という個別の状況には，考慮するべき点があると考えられる．単純な重回帰分析では比較的見逃されやすい個別データの特徴を抽出できる点が，注意機構を導入する第一の利点と言える．

第二は多重共線性の問題である．重回帰分析では前処理段階で多重共線性を起こしているデータを分析者が調整することが多く見られる．理論的な背景や推定モデルがない場合，変数の削除は分析者の力量に依存する．注意ベクトルを用いる方法において，回帰式の基

項目 1	項目 3		項目 1	項目 3
1	3		3	4
1	3		3	5
1	3		4	1
1	3		4	2
1	3		4	3
2	1		4	4
2	2		4	5
2	3		5	1
2	4		5	2
2	5		5	3
3	1		5	4
3	2		5	5
3	3			

表 1: アンケート結果の具体例

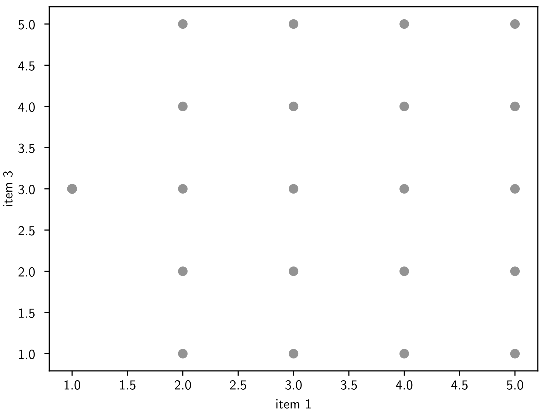

図 2: 表 1 の散布図

本形は，

$$y = \sum_{j=1}^{l} w_j m_j b_j(x_j) + c$$

となっている．これより，注意の数値 m_k が 0 に近づくと，説明変数 x_k の係数 w_k も 0 になる．

つまり，機械学習の効果によって，説明変数間の多重共線性の問題が学習され，また自動調整される．前処理をすることなく重回帰モデルが利用できる点が，第二の特徴と言える．多重共線性を回避する手段もいくつか知られているが，本稿の注意機構付き回帰モデルは，変数の影響度を判定することが本来の主目的であり，影響度が低くなることで自動的に多重共線性の問題も回避される点は大きな違いとなっている．

第三は，解釈可能な非線形重回帰モデルを構築できる点である．通常のニューラルネットワーク同様，全結合層 1 層目の出力のノード数を複数に変更し，全結合層・活性化関数を多層にするモデルの精度を上昇させることが可能である．注意ベクトル部分の影響により，影響力のない変数については，0 へ近づき，逆に，影響力のある変数については大きな確率が自動的に割り当てられることとなる．この解釈可能な非線形モデルの構築が，第三の特徴と言える．

3　データ解析の実例

本稿では，scikitlearn から入手できるボストン住宅価格データを利用して，重回帰モデルと注意機構付き重回帰モデルを比較しつつ，注意ベクトルの特徴について検討する．ボストン住宅データは，506 個のデータから構成されており，そのすべてを分析データとして利用した．[*1]

重回帰モデルの基本的結果を述べる．表 2 より，5 番目の RM の住居平均部屋数の影響力が大きいことがわかる．3 番目の CHAS は，ダミー変数で，プラスの効果を持つ．[*2]また，12 番目の LSTAT は，マイナスの効果を持っている．一方，係数が 0 に近い，6 番目の AGE，11 番目の B に関しては，影響力がほとんどなく，また，2 番目の INDUS，9 番目の TAX などは，比較的影響力が少ないと言える．NOX や DIS については，マイナスの影響が大きい．決定係数は，0.74 である．

図 3 のヒートマップは，色が濃い程，相関係数が 1 に近く，逆に，白いほど相関係数が −1 に近くなる．相関行列の数値より，RAD と TAX には強い相関が観測される．このことから，RAD と TAX については，多重共線性が起こっているとみなされる．

次に，注意機構付き回帰分析の結果を，全結合層が 2 層の非線形モデルと線形モデルにわけて考察する．図 4 と図 5 のグラフは，50 回の平均値によるグラフであり，学習回数は，20000 回となっている．

[*1] 各変数の意味合いを簡単に説明する．CRIM (犯罪発生率)，ZN (住居区画の密集度)，INDUS (非小売業の土地割合)，CHAS (川の周辺かどうか)，NOX (NOx 濃度)，RM (住居の平均部屋数)，AGE (物件の年代割合)，DIS (雇用施設からの重み付き距離)，RAD (大きな道路へのアクセス)，TAX (所得税率)，PTRATIO (教師あたりの生徒数)，B (黒人の比率)，LSTAT (低所得者の割合) とされる．被説明変数は，住宅価格となる．

[*2] Utsumi (2019) ダミー変数を埋め込みベクトル化して効率的に回帰モデルに取り込むことも可能である．

	Features	Coefficient Estimate
0	CRIM	-0.108011
1	ZN	0.046420
2	INDUS	0.020559
3	CHAS	2.686734
4	NOX	-17.766611
5	RM	3.809865
6	AGE	0.000692
7	DIS	-1.475567
8	RAD	0.306049
9	TAX	-0.012335
10	PTRATIO	-0.952747
11	B	0.009312
12	LSTAT	-0.524758
	二乗誤差	決定係数
	21.9	0.74

表 2: 重回帰モデルでの係数と基本データ

	CR	ZN	IN	CH	NO	RM	AG	DI	RA	TA	PT	B	LS
CR	1	-0.2	0.4	-0.1	0.4	-0.2	0.4	-0.4	0.6	0.6	0.3	-0.4	0.5
ZN	-0.2	1	-0.5	0	-0.5	0.3	-0.6	0.7	-0.3	-0.3	-0.4	0.2	-0.4
IN	0.4	-0.5	1	0.1	0.8	-0.4	0.6	-0.7	0.6	0.7	0.4	-0.4	0.6
CH	-0.1	0	0.1	1	0.1	0.1	0.1	-0.1	0	0	-0.1	0.1	-0.1
NO	0.4	-0.5	0.8	0.1	1	-0.3	0.7	-0.8	0.6	0.7	0.2	-0.4	0.6
RM	-0.2	0.3	-0.4	0.1	-0.3	1	-0.2	0.2	-0.2	-0.3	-0.4	0.1	-0.6
AG	0.4	-0.6	0.6	0.1	0.7	-0.2	1	-0.8	0.5	0.5	0.3	-0.3	0.6
DI	-0.4	0.7	-0.7	-0.1	-0.8	0.2	-0.8	1	-0.5	-0.5	-0.2	0.3	-0
RA	0.6	-0.3	0.6	0	0.6	-0.2	0.5	-0.5	1	0.9	0.5	-0.4	0.5
TA	0.6	-0.3	0.7	0	0.7	-0.3	0.5	-0.5	0.9	1	0.5	-0.4	0.5
PT	0.3	-0.4	0.4	-0.1	0.2	-0.4	0.3	-0.2	0.5	0.5	1	-0.2	0
B	-0.4	0.2	-0.4	0.1	-0.4	0.1	-0.3	0.3	-0.4	-0.4	-0.2	1	-0.4
LS	0.5	-0.4	0.6	-0.1	0.6	-0.6	0.6	-0.5	0.5	0.5	0.4	-0.4	1

表 3: 相関行列の表，変数名は，先頭 2 文字のみで表示してある．

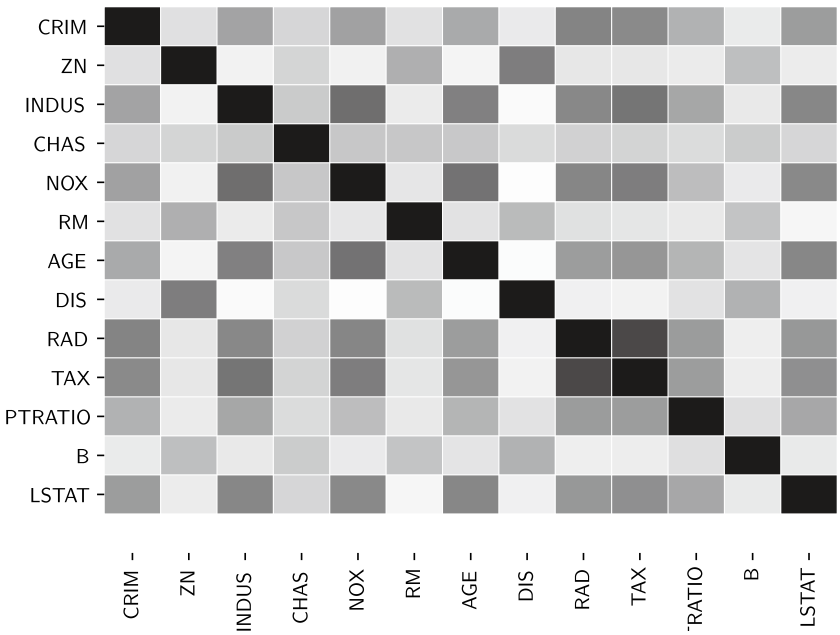

図 3: 相関行列のヒートマップ

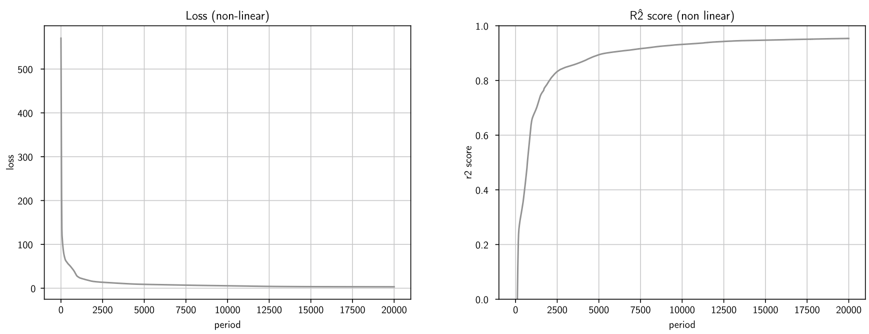

図 4: 非線形モデルの損失（左）と決定係数（右）の推移

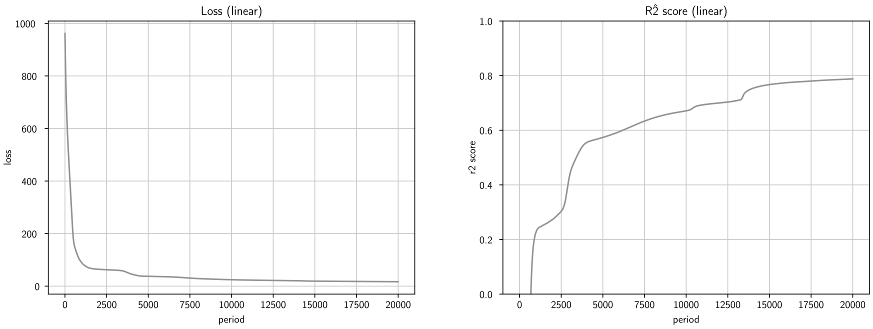

図 5: 線形モデルの損失（左）と決定係数（右）の推移

最初に，非線形モデルから考察する．損失は，学習回数が 2500 回前後まで，急減少している．決定係数に関しては，初期のモデルの適合性は悪く，マイナスの値から始まっている．こちらも，2500 回ほどで 0.85 前後に到達する．20000 回の繰り返し計算によって，非線形モデルの最小 2 乗誤差は，ほぼ 3.5 前後の間に，また，決定係数は，0.95 前後に収まった．15000 回以降は，決定係数の上昇幅は小さく，また，損失の減り方も非常に緩やかであった．

表 4 は，非線形モデルの注意ベクトルの重みに関する基本統計量である．個別のデータの注意ベクトルの重みから，平均値，標準偏差，最小値，最大値を求めたものである．表 4 より，RM，NOX，LSTAT が大きな影響力を持っていることがわかる．いずれの変数も，0.00，少数第 2 位まで 0 と，標準偏差が非常に小さい．このことより，平均値の影響力がモデル全体の影響力とみなせる．

個別データから判断すると，若干，異なる様相を呈している．住宅価格に影響力を与えている上位 3 変数は，RM，NOX，DIS となっており，全体の平均としては，LSTAT が有効な変数と思われているが，個別データとしては，それほど影響力がない事がわかる．実際，LSTAT は，3 番目の影響力に初めて登場しており，その割合は，0.0097 である．全体としては，影響力があるように見えるが，個別データではそれほど影響力がない具体例となっている．DIS は，全体としては影響力は小さいが，個別データにおいては，大きな影響力を持っている具体例となっている．DIS は，影響力の 2 番目，3 番目の出現回数の合計は，413 であり，頻度としては，0.816 となる．約 8 割の個別データについて，DIS の強い影響を受けていることが観察される．

影響力がほとんどないとみなされる変数は，AGE，CRIM，B，TAX の 4 種類である．このあたりは，重回帰モデルの影響力と同じと考えられる．また，TAX の注意ベクトルの重みが 0 に近づいていることがから，TAX と RAD の強相関に関しては，自動的に学習され，解消されていることがわかる．

表 7 は，利用したデータの数値と影響度の関係を表現した表である．濃淡により影響度が表現され，濃い程，影響度が高い．例 1 や例 2 は，モデルの中の 95%で観測される影響度のパターンである．一方，例 3 や例 4 は，ZU が影響を与えるデータである．ZU の値が影響力の要因になっている事がわかる．モデル全体としては影響力は小さいが，個別デー

タにおいては，大きな影響力を持っている個別データが存在することがわかる．

	mean	std	min	max
CRIM	0.011985	0.001141	0.004734	0.013972
ZN	0.06965	0.043263	0.040642	0.283705
INDUS	0.062416	0.003606	0.04715	0.069336
CHAS	0.051061	0.002701	0.038293	0.055883
NOX	0.15079	0.007664	0.117434	0.169843
RM	0.240592	0.020089	0.172161	0.300482
AGE	0.010357	0.003554	0.003785	0.022443
DIS	0.141472	0.025555	0.077509	0.204181
RAD	0.033359	0.003035	0.026564	0.039634
TAX	0.037713	0.007806	0.021511	0.058788
PTRATIO	0.0541	0.007535	0.035012	0.080687
B	0.028572	0.003498	0.020573	0.035903
LSTAT	0.107935	0.016957	0.055214	0.147101

表 4: 非線形モデルの基本統計量

影響力	1	2	3	4
第 1 位	RM: 488 (0.964)	ZN: 18 (0.035)		
第 2 位	NOX: 287 (0.567)	DIS: 183 (0.433)	ZN: 18 (0.035)	RM: 18 (0.035)
第 3 位	DIS: 230 (0.455)	NOX: 219 (0.433)	LSTAT: 49 (0.097)	ZN: 8 (0.016)

表 5: 非線形モデルの　影響力上位 3 の変数の数と割合

全結合層が 1 層の線形モデルを考察する．通常の重回帰に近いモデルなので，決定係数や二乗誤差も近い値を示している．最小 2 乗誤差は，17，決定係数は，0.78 に近い値を示した．線形モデルは，重回帰モデルと近い結論に達している．

表 8 は，線形モデルの注意ベクトルの重みに関する基本統計量である．非線形モデル同様，個別のデータの注意ベクトルの重みから，平均値，標準偏差，最小値，最大値を求めたものである．表 8 より，モデル全体としては，RM の影響力が 0.55 と突出している．また，LSTAT も 0.178 と，約 18%の影響力を持っていることがわかる．RM，LSTAT，INDUS を除けば，標準偏差は，0.00，少数第 2 位まで 0 であり，重みにの数値に関しては，幅がない事がわかる．一方，RM は，最小 0.1，最大 0.725 と影響力に大きな幅がある．住宅価格の約 7 割を RM で説明できるデータもあれば，1 割の状況もある．

個別データから判断すると，若干，異なる様相を呈している．住宅価格に影響力を与えている上位 3 変数は，RM，LSTAT，PTRATIO となっており，モデルの平均的な影響力と変わらない．しかし，PTRATIO は，3 番目の影響力だが，その割合は，8 割を超えるもので

影響力	1	2	3
第 13 位	AGE: 359 (0.709)	CRIM: 147 (0.291)	
第 12 位	CRIM: 359 (0.709)	AGE: 145 (0.287)	B: 2 (0.004)
第 11 位	B: 427 (0.843)	TAX: 77 (0.152)	AGE: 2 (0.004)

表 6: 非線形モデルの影響力下位 3 の変数の数と割合

	例 1	例 2	例 3	例 4
PRICE	24.0	33.2	35.400	30.300
CRIM	0.006	0.104	0.013	0.0466
ZN	18.0	40.0	90.0	80.0
INDUS	2.309	6.409	1.220	1.519
CHAS	0	1	0	0
NOX	0.537	0.446	0.402	0.404
RM	6.574	7.26	7.249	7.106
AGE	65.199	49.0	21.899	36.599
DIS	4.090	4.787	8.696	7.309
RAD	1	4	5	2
TAX	296.0	254.0	226.0	329.0
PTRATIO	15.300	17.60	17.899	12.600
B	396.899	389.2	395.929	354.309
LSTAT	4.980	6.050	4.809	8.609

表 7: 非線形モデルにおける PRICE への影響度が大きい上位 4 変数の影響度

あり，効果は小さいが，多くのデータはっきりと影響を与えていることがわかる．基本的に線形タイプなので，個別の状態がそのまま全体の状況に伝播している．

影響力がほとんどないとみなされる変数は，NOX，CHAS，RAD，CRIM，ZN などである．このあたりは，重回帰モデルの影響力と同じと考えられる．また，RAD の注意ベクトルの重みが 0 に近づいていることがから，TAX と RAD の強相関に関しては，自動的に学習され，解消されていることがわかる．非線形モデルでは，影響力が大きい NOX の影響が極めて少なく評価されている点が重要である．線形と非線形の効果をわけている部分と見なすことができる．ダミー変数の CHAS の影響が完全に取り除かれている．

注意機構付き重回帰分析を行うことで，全体としては，影響力があるように見えるが，個別データではそれほど影響力がない個別データの存在が明らかになる．同様に，全体としては影響力は小さいが，個別データにおいては，大きな影響力を持っている個別データの存在も明らかになる．

4　帰結

注意機構付き重回帰分析を行うことで，全体としては，影響力があるように見えるが，個別データではそれほど影響力がない個別データの存在が確認された．同様に，全体としては影響力は小さいが，個別データにおいては，大きな影響力を持っている個別データの存在も確認された．つまり，モデル全体での影響力と個別データごとには，変数間の影響力の違いが見られる．個別データでの影響力を求められることが，機械学習を利用した重回帰モデルの特徴と言える．具体的には，個別データ毎に，影響力のある変数を求めることができるため，そのデータに応じた予測や対応なども可能となる．

	mean	std	min	max
CRIM	0.000412	0.000315	0.000102	0.004758
ZN	0.001469	0.003483	0.00012	0.026014
INDUS	0.044204	0.132805	1.7E-05	0.715192
CHAS	4.945428E-06	6.706003E-07	1.508473E-06	6.072377E-06
NOX	1.709193E-06	2.771535E-07	5.082219E-07	2.522723E-06
RM	0.550074	0.100435	0.103744	0.725062
AGE	0.019778	0.018931	0.000919	0.087142
DIS	0.060796	0.016693	0.015229	0.114947
RAD	0.002434	0.005652	3.7E-05	0.025476
TAX	0.026701	0.009499	0.004369	0.050137
PTRATIO	0.08946	0.030961	0.038371	0.302233
B	0.025821	0.006231	0.003021	0.036841
LSTAT	0.178843	0.054318	0.016773	0.294469

表 8: 線形モデルの基本統計量

影響力	1	2	3	4	5
第 1 位	RM: 475 (0.939)	INDUS: 31 (0.061)			
第 2 位	LSTAT: 445 (0.879)	RM: 31 (0.061)	PTRATIO: 18 (0.035)	INDUS: 12 (0.024)	
第 3 位	PTRATIO: 440 (0.870)	DIS: 28 (0.055)	LSTAT: 23 (0.045)	AGE: 8 8 (0.016)	INDUS: 7 (0.014)

表 9: 線形モデルの影響力上位 3 の変数の数と割合

影響力	1	2	3	4
第 13 位	NOX: 506 (1)			
第 12 位	CHAS: 506 (1)			
第 11 位	RAD: 211 (0.417)	CRIM: 114 (0.225)	ZN: 108 (0.213)	INDUS: 73 (0.144)

表 10: 線形モデルの影響力下位 3 の変数の数と割合

参考文献

[1] J., Devlin, M., Chang, K., Lee and K., Toutanova (2018) “BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding”, NAACL-HLT

[2] H. Fukuni, Hirakawa, Yamashita, Fujiyoshi (2019) “Attention Branch Network: Learning of Attention Mechanism for Visual Explanation”, Computer Vision and Pattern Recognition, 10705–10714

[3] Schutze, Xiang, and Zhou. (2015) “Attention-Based Convolutional Neural Network for Modeling Sentence Pairs.” CoRR, abs/1512.05193

[4] Y. Utsumi (2019) “Notes on Quantification Theory I Using a Neural Network Approach,” The Journal of Chiba University of Commerce, Vol. 57, No. 2, 85–93

[5] Vaswani, Shazeer, Parmar, Uszkoreit,Jones, Gomez, Kaiser, and Polosukhin. (2017) “Attention Is All You Need”, Advances in Neural Information Pro-cessing Systems, 6000-6010.

（2020.8.26 受稿，2020.10.27 受理）

〔抄　録〕

本稿の目的は，自然言語処理で近年利用されている，注意機構を導入した回帰分析を提案することである．注意機構を導入することによって，回帰係数の影響力を割合として表しつつ，かつ，個別のデータの変数の影響度も同時に知ることが可能となる．注意機構付き回帰分析を行うことで，全体としては，影響力があるように見えるが，個別データではそれほど影響力がない個別データの存在が，本稿では確認された．また，同様に，全体としては影響力は小さいが，個別データにおいては，大きな影響力を持っている個別データの存在も確認された．個別データでの影響力を求められることが，機械学習を利用した回帰モデルの特徴と言える．これにより，個別データ毎に，影響力のある変数を求めることができるため，個別のデータに応じた予測が可能となる．

―Abstract―

The purpose of this paper is to propose a regression analysis with an attention mechanism, which has recently been used in natural language processing. By introducing an attention mechanism, it is possible to express the influence of the regression coefficients as a percentage, while at the same time knowing the influence of the individual data. This paper confirms the existence of individual data that appear to be influential in the model, but not so influential in the individual data, through regression analysis with an attention mechanism. Similarly, some individual data have a small influence on the whole, but have a large influence on individual data. This makes it possible to obtain influential variables for each individual data, and thus to make predictions according to the individual data.