

二輪自動車における新しい 信用リスクモデルの提案

東海大学 情報通信学部 経営システム工学科

藤沼 亮太

目次

1. 研究背景

2. データ概要

3. 信用リスクモデルの評価方法

4. 分析手順

5. 分析概要

6. 分析結果

7. 考察

8. 今後の課題

参考文献

Appendix

1. 研究背景 ~近年の動向~

日本

規制緩和、技術革新等で日本企業の経済環境は激変し、バブル崩壊以降の日本経済の低迷も影響してデフォルト^[AP.I]件数は増加傾向にある。

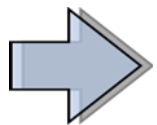
アメリカ

サブプライムローン^[AP.II]問題により世界中の投資コミュニティに影響を与えた。

ヨーロッパ

信用リスク関連の問題のために、ドイツの数行の銀行は破綻、吸収された。イギリスでは取付騒動により、困難に陥った住宅ローンの貸し手である銀行が国有化される結果となった。

信用リスク^[AP.III]に応じた収益確保が新BIS規制^[AP.IV]に向け**世界で重要な課題**に



しかし中南米国では**ローンに対する認識が甘い**

1. 研究背景 ~中南米国のローン問題~

□ 中南米国政府が実際に行った政策

- ✓ 公的な金融機関による貸出枠を拡大し、利子を下げる。
- ✓ 中小企業への融資を公的責任で行う。

⇒ 貧困層

比較的ローンを組みやすくなったが、
ローン経験の不足からローンの返済忘れ
が多発するようになる。

- ## □ ストラテジックデフォルトという、 支払い能力があるにも関わらずローン返済 をしない顧客の発生

⇒ 富裕層

富裕層にもローンに対する意識が甘い顧客がいる。

◆ 政策金利の推移

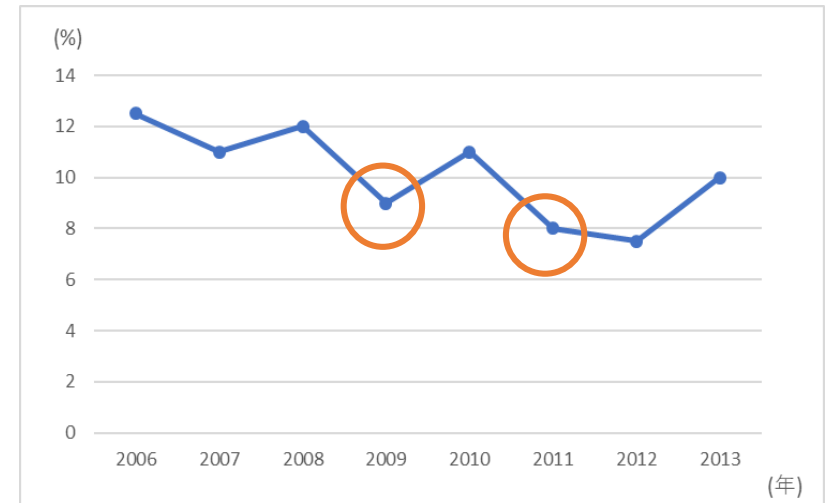


図1 中南米の政策金利の推移

1. 研究背景 ~現状と問題点~

現状

デフォルト確率を正確に予測し、**信用リスクに見合うだけの収益を確保する**という考え方が**重要**だということが知られている。しかしながら現在の中南米国ではローンの踏み倒しが多発しており**企業が製造費を回収できず、収益を確保できていないことが現状**である。



研究目的

本研究ではどのような**信用リスクモデル**^[AP.V]の精度が高いのかを比較する。また**デフォルトしやすい人物の特徴、デフォルトしない人物の特徴**を把握する。これは企業にとって有益であり、**ローンの問題を改善しやすくし、収益の確保が果たせる**と考えられる。

2. データ概要

二輪自動車販売ローンに関する顧客データ

表1 データ詳細

提供	二輪自動車販売会社
期間	2010年9月1日～2012年6月30日
エリア	中南米

データ概要

データ数：13214件

欠損値削除後：13123件

データ変数：22変数

変数内容

在住、学歴、職業、結婚歴、就労年数、購入した商品名など
個人の属性を表すデータとローン返済の有無

変数詳細

本研究で使用した変数は0,1の名義尺度と
連続データの間隔尺度である。

3. 信用リスクモデルの評価方法

評価方法の特性について**判別力(Discriminative Power)**と予測と結果の**合致性(Calibration Power)**の2つの側面があると知られている。

- ✓ **判別力**：予測デフォルト確率が良く当たる、的中率の高いモデル
Ex)事前に想定されたモデルの的中率が60%で予測結果が90%的中になる
- ✓ **合致性**：モデルには誤差があるため予測結果も誤差を含んでいる
よって予測と予測結果が合致するモデルがよいモデルである
Ex)事前に想定されたモデルの的中率が60%で予測結果も60%的中となる

運用する立場にとっては、損失を最小限に止めるために、
予測結果の的中率が高いモデル(**判別力**)を用いるはずである

そのため本研究では**精度が高いものを良いモデルとする**

※本研究で扱う信用リスクモデルは貸し倒れではなく未払いの顧客を対象に研究する。

3. 信用リスクモデルの評価方法

従来の信用リスク評価における統計モデル

✓ 判別分析

✓ ロジットモデル

✓ 機械学習



説明力の高い統計モデル

統計モデルの中でも機械学習の一つであるニューラルネットワーク分析は高い説明力を持っていることが知られている。しかし、ニューラルネットワークのようなブラックボックスであるモデルは運用者が**推論過程**において入出力間の関係を理解しづらいという**デメリット**が存在する。



本研究で使用する統計モデル

近年、脚光を浴びるニューラルネットワークと説明変数をスコア化でき、多数の金融機関の内部モデルでも採用されている手法であることから、**ロジスティック回帰分析**で**精度の比較**を行う。

4. 分析手順

分析

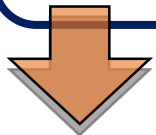
ロジスティック回帰分析

P値を求め、回帰式に有意な説明変数を探る。
オッズ比を出すことにより、どの説明変数が
デフォルトする人物、デフォルトしない人物に影響するのか探る。



ディープラーニング

ロジスティック回帰分析と同じ説明変数を用いて学習させ
精度を評価する。



精度比較

精度を比較し、どちらのモデルが良いのかを検討する。

5.1 分析概要 ~ロジスティック回帰分析~

- ロジスティック回帰分析とは質的変数（2値データ）を目的変数とするときに用いられる回帰分析である。
- この解析手法は2群の判別問題や発生確率の予測問題に適用される。
- 評価プロセス
 - ✓ モデルに含まれる説明変数の有意性を評価するプロセス
 - ✓ モデル全体の適合度を評価するプロセス

本研究ではどちらの評価も行う

- 本研究ではVRPを用いる。VRPならば複雑なプログラミングをいかずとも簡単にロジスティック回帰分析が行える。
- データクリーニングや変数の属性変更に関してはVMSを使用した。



図2 ロジスティック回帰分析のフロー図

5.1 分析概要 ~ロジスティック回帰分析~

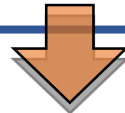
目的変数

本研究では信用リスクモデルによりデフォルト・非デフォルトを見分ける必要があるため目的変数にはデフォルト・非デフォルトの2値を設定する。

説明変数

自動二輪車購入日など明らかに説明変数にそぐわ変数や欠損値が多すぎる変数は削除した。

また、回帰分析を行う上で**多重共線性**を考慮しなくてはならない。



本研究で用いた変数は間隔尺度と名義尺度の2つであった。

- ✓ 間隔尺度+間隔尺度：積率相関係数 (0.33未満で相関なしとする)^[13]
- ✓ 名義尺度+名義尺度：連関係数 (0.1未満で相関なしとする)^[14]
- ✓ 間隔尺度+名義尺度：相関比 (0.2未満で相関なしとする)^[15]

相関があった変数同士に関しては片方の変数を削除する。

5.1 分析概要 ~ロジスティック回帰分析~

□ 多重共線性を考慮した変数を以下に示す。

- ✓ ネガティブ情報
- ✓ 性別
- ✓ 在住
- ✓ 就労年数
- ✓ 収入
- ✓ 年齢
- ✓ 与信限度額
- ✓ 返済能力
- ✓ ローン分割数
- ✓ ローン会社からの照会件数

□ P値の分析結果を以下に示す。

P値は回帰式にどれだけ優位に関わっているかという統計学指標として用いられる。一般的に**有意水準0.05**^[16]で回帰式に影響していると言えるため、P値が0.05以上の変数を表2に示した。

◆ P値が0.05以上の変数

表2 ロジスティック回帰分析結果

変数名	p値 ≥ 0.05
収入	0.19
年齢	0.19
与信限度額	0.75
返済能力	0.76

よってロジスティック回帰分析にはP値0.05未満の**6つ**の変数を用いる

5.1 分析概要 ~ロジスティック回帰分析~

- オッズ比による説明変数の比較
オッズとは事象が起こる確率と起こらない確率の比のことである。
たとえばある事象が発生する確率をPとすると、オッズは $P/(1-P)$ となる。
さらにこのオッズとオッズの比をオッズ比と呼ぶ。オッズに差がなければオッズ比は1となる。
- 本研究では1より大きい数字ほどデフォルト人物になる確率への影響力が大きくなる。0に近づくほど非デフォルト人物になる確率への影響力が大きい。

◆ オッズ比での変数の比較

表3 オッズ比の結果

	オッズ比
ネガティブ情報	0.066
性別	0.719
在住	1.392
就労年数	1.178
分割数	0.973
ローン会社からの照会件数	1.024

表3から**在住、就労年数、ローン会社からの照会件数**がデフォルトの人物に影響する変数であり、**ネガティブ情報、性別、分割数**が非デフォルト人物に影響する変数である。

5.1 分析概要 ~ロジスティック回帰分析~

- モデル全体の適合度を評価するプロセスとして**判別精度**を見て評価する。
- 判別精度を見るためにVRP機能の「**予測**」を用いた。

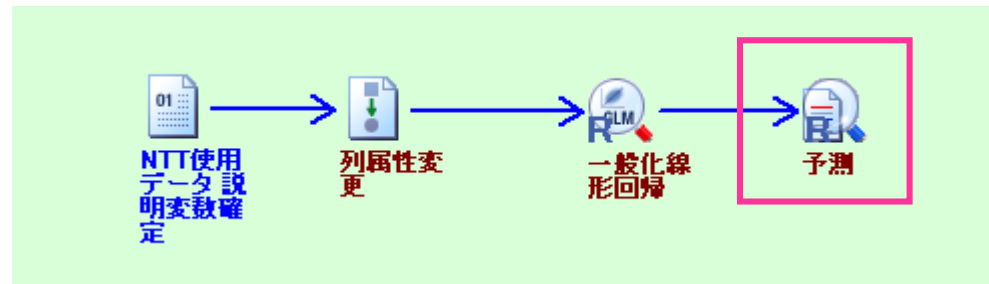


図3 判別精度予測フロー図

- 判別精度の結果を以下に示す。

判別の正解した数は表4の
(0,Prediction.X0),(1,Prediction.X1)
を見れば良い。

◆ 判別精度の結果

表4 「予測」を使った判別精度

合否	Prediction.X0	Prediction.X1	Sum
0	10075	96	10171
1	3007	36	3034
Sum	13082	132	13214

5.2 分析概要 ~ディープラーニング~

- 本研究ではDeep Learnerを使用し、機械学習を行った。
Deep LearnerではVRP同様、プログラミングの知識が無くてもディープラーニングを実行することができる。
- ロジスティック回帰分析のときと同様にデータクリーニングにはVMSを使用した。
- 目的変数にはデフォルト、非デフォルトの2値を設定し、説明変数にはロジスティック回帰分析と同様に6変数を用いる。
- Deep Learnerの学習については、13123件のデータのうち、ランダムで8割(10498件)を学習用に、2割(2625件)を検証用に使用した。

➡ 分け方についてはVMSの「**ファイル分割**」を用いている

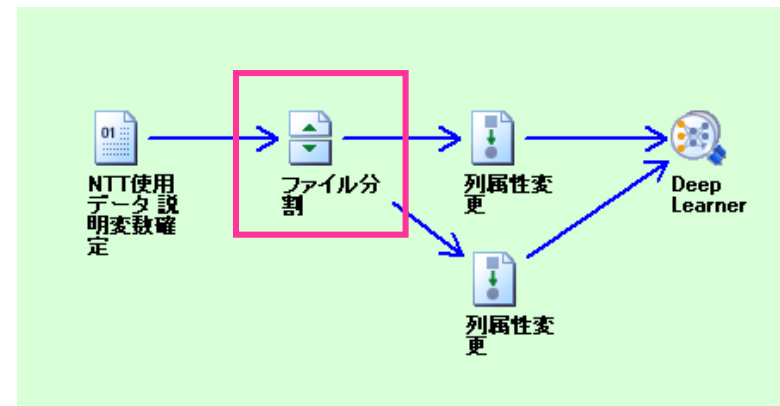


図4 ディープラーニングのフロー図

5.2 分析概要 ~ディープラーニング~

- 本研究で作成するディープラーニングモデルは予測対象がデフォルトする顧客とデフォルトしない顧客の2値カテゴリの**分類モデル**である。
- ディープラーニングはニューラルネットワークの一種であるため、信用リスクモデルを作成するにあたって問題はないと判断した。

ニューラルネットワークの構造は
多層パーセプトロンという
ネットワーク構築をもつため
本研究でもこの構築を用いる

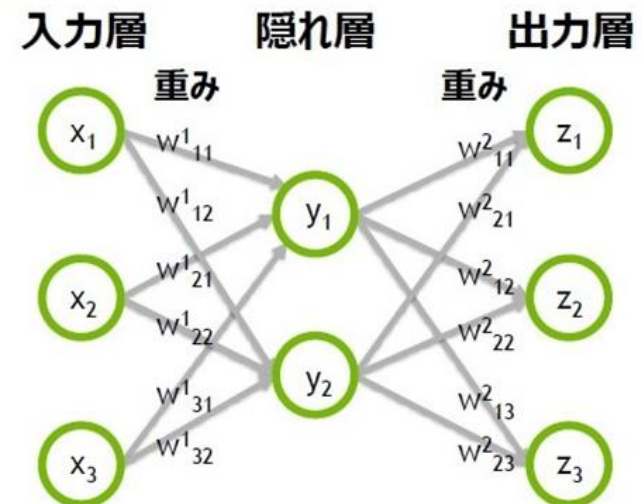
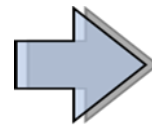


図5 多層パーセプトロン

5.2 分析概要 ~ディープラーニング~

□ Deep Learnerでは中間層の数を任意で指定することが出来る

中間層の層数の決定法

中間層の数は任意で設定することができるが、少なくとも2つの中間層が必要であると言われている。また、隠れ層の数が多くなるほど確率的勾配降下法では勾配消失問題、勾配の不安定化が生じると言われているため、中間層の数を増やしすぎてもいけない。

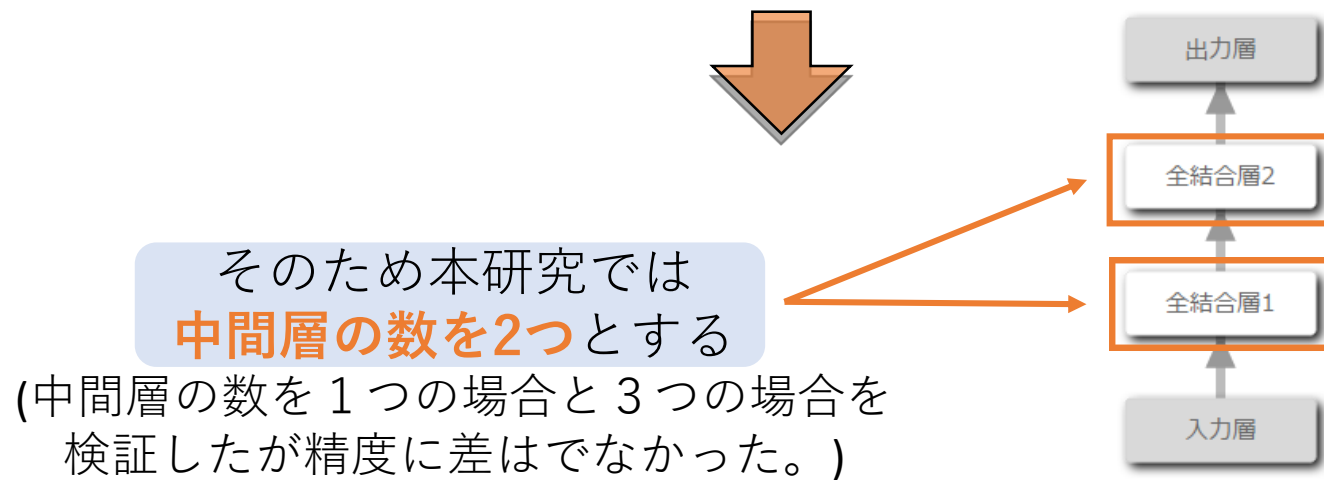


図6 中間層の数

5.2 分析概要 ~ディープラーニング~

- 中間層では、
 - ✓ **出力次元数**
 - ✓ **活性化関数**
 - ✓ **Dropout Ratio**の3つのパラメーターを任意で選択することができる。

全結合層2

Model Optimizer 探索範囲

出力次元数

最小値: 1 最大値: 100

活性化関数

ReLU sigmoid tanh softmax 線形

Dropout Ratio

最小値: 0.0 最大値: 0.5

図7 中間層のパラメーター

Deep Learnerではこれら3つのパラメーターを自動で調節することが可能な**Model Optimizer**が備わっていたため、本研究ではこの機能を用いる。

出力次元数	活性化関数	Dropout Ratio

図8 Model Optimizerの表示

5.2 分析概要 ~ディープラーニング~

- Model Optimizerの設定についても任意で設定することができる
Model Optimizerの設定

手法： Random Forest を利用したベイズ最適化によりパラメータを探索する「**ベイズ最適化**」を選択肢した

指標：モデルを評価する際の評価指標を以下から選択することができる

✓ Loss ✓ F1-score ✓ Recall ✓ Precision ✓ Accuracy

➡ 企業はデフォルト人物を見抜けないと損失し、非デフォルト人物を見抜けないと貸付機会を失い損失する。そのため**Accuracy**を選択する。

最大探索数：モデルを評価する最大回数を指定する

➡ 機械学習は毎回出力される結果に誤差がでる。本研究では**5回**評価し、その平均を精度とした。

使用プロセス数：パラメータ探索のために使用するプロセス数を指定する

➡ 本研究ではデフォルトの**1**とした。

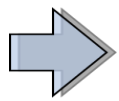
(※ Deep Learner技術資料参照)

5.2 分析概要 ~ディープラーニング~

- 学習についても任意で設定することができる

学習設定

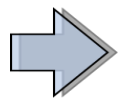
ミニバッチサイズ：モデルのパラメータ更新をするまでに与えるデータサイズを指定する



通常 2^n が良いとされているため初期値の64と1024を試したが精度に差はでなかったため初期値である、**64**を採用する。

SGD設定：SGD 最適化に使われるアルゴリズムを以下から選択する

✓ Adam ✓ RMSProp ✓ AdaGrad ✓ AdaDelta ✓ SimpleSGD



SGD設定は最適なものを使用してよいため全ての設定を検証し、**一番精度の高いもの**を採用する。

(※ Deep Learner技術資料参照)

5.2 分析概要 ~ディープラーニング~

- 学習についても任意で設定することができる

学習設定

エポック数：学習のためにデータ全体を何回繰り返して与えるかを指定する

➡ エポック数は過学習を起こさず、訓練精度と予測精度がともに高いというものが良いとされる。そのため学習がある程度進んだ段階で打ち切る方法をとる。エポック数は**100**とする。

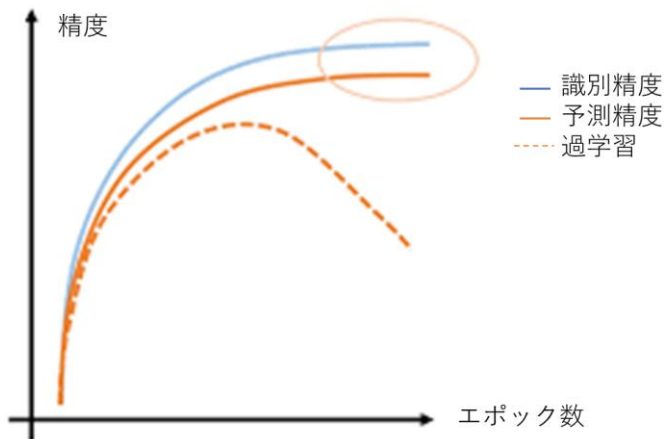


図9 エポック数と精度の関係

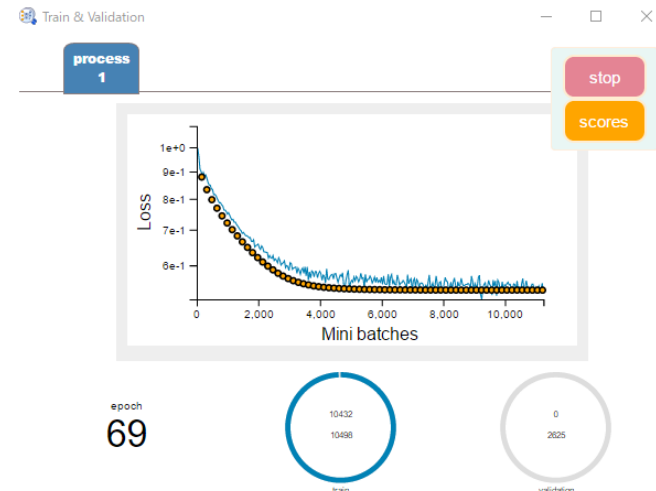


図10 Deep Learner学習の様子

(※ Deep Learner技術資料参照)

5.2 分析概要 ~ディープラーニング~

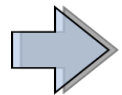
- 学習についても任意で設定することができる

学習設定

学習率：学習時にパラメータを更新する大きさを指定する

学習率の決め方に「LR range test」というものがある。学習率を徐々に増加させながら機械学習の学習率を決める手法である。

10^{-7} から学習率を10倍ずつ試し、1番精度の良いものを採用する。



学習率は 10^{-4} から精度が変わることはなかったため本研究では **10^{-4}** を使用する。

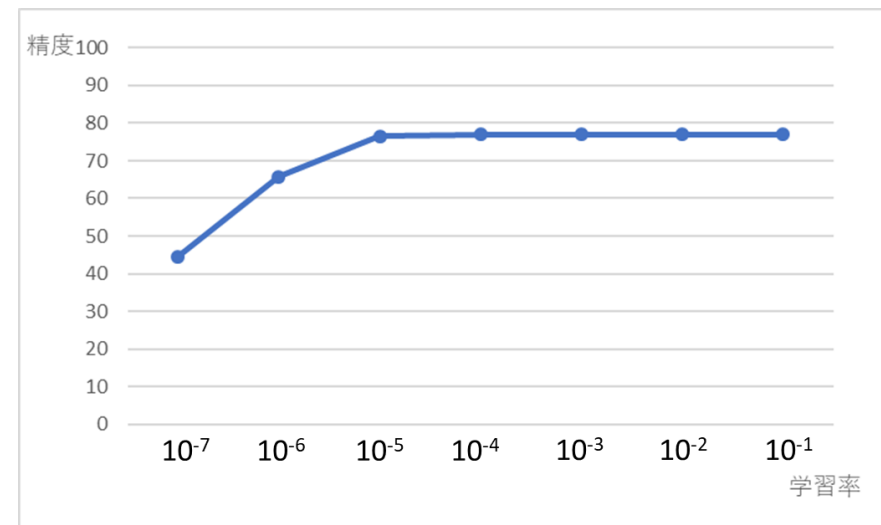


図11 学習率と精度の関係

(※ Deep Learner技術資料参照)

6. 分析結果

- ロジスティック回帰分析についてもディープラーニング同様、Accuracyでの正解率を出して比較する。
- Accuracyの数値は(1)式で算出することができる。

$$Accuracy = \frac{True\ Positive + True\ Negative}{True\ Positive + False\ Positive + False\ Negative + True\ Negative} \quad (1)$$

ロジスティック回帰分析の精度は **Accuracy= 0.7652**

- ディープラーニングのAccuracy精度を表5に示す。

◆ ディープラーニング判別精度

表5 機械学習精度(Accuracy)

	Adam	SimpleSGD	AdaDelta	AdaGrad	RMSProp
Accuracy	0.770	0.780	0.780	0.668	0.780

ディープラーニングの精度は **Accuracy= 0.780**

7. 考察

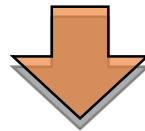
本研究から分かること

- ロジスティック回帰分析とディープラーニングのAccuracyによる精度の比較については、**ディープラーニングの方が精度が高い**結果となった。
- ロジスティック回帰分析ではオッズ比による説明変数の比較により**デフォルトする人物の特徴は在住**によく表れており、**デフォルトしない人物の特徴はネガティブ情報**によく表れているということが分かった。
- ネガティブ情報は他の変数に比べ、非デフォルトに大きく偏っている。やはり、**直接ローンに関わる変数が重要であると言える。**
- 今回の研究背景で貧困層、富裕層ともにデフォルトする人物がいると述べたが**就労年数(収入)**はオッズ比から見ると、**デフォルト人物の特徴である変数**だということが分かった。

7. 考察

企業のメリット

- 本研究で作成したモデルではディープラーニングを用いたモデルにより高い精度が得られた。これにより**損失機会が大きく削減できると考えられる。**
- 前述したとおりディープラーニングによるモデルはブラックボックスモデルであるため、どのような変数がモデルに影響を及ぼしているのか分からず、企業にとってはデメリットである。精度ではなく、**変数の影響を見る意味でロジスティック回帰分析の活用が考えられるだろう。**



- **企業の損失機会を最小限にすることを意識した本研究のモデルは企業の収益確保に大きく貢献すると考えられる。**

8. 今後の課題

- 信用リスクモデルにおいてモデルの精度向上は大きな課題である。本研究ではロジスティック回帰分析と同様の説明変数を用いて機械学習させたがより良い説明変数を見つけていきたい。
- また、機械学習の代表的な手法として本研究で行ったNeural Networkの他に Gradient Boosting Decision Tree , Random Forest , Support Vector Machineがあり、どの手法が最適かは適用する分野やデータセットの特性に依存する。機械学習では一般的に単一モデルのみを学習させるよりも複数のモデルを組み合わせるアンサンブル学習の方が高い精度が得られる場合が多いことが知られている。本研究で使用したデータに最も合う分析手法の発見が今後の大きな課題である。

参考文献

- [1]安道知寛,山下智志：財務指標の時間依存を考慮した信用リスク評価モデル-デフォルト予測への応用,(2004)
- [2]澤木太郎,田中拓哉,笠原亮介：機械学習による中小企業の信用スコアリングモデルの構築,人工知能学会研究資料
- [3]山下智志,川口昇,敦賀智裕：信用リスクモデルの評価方法に関する考察と比較, *Financial Research and Training Center discussion paper series*, 11, (2003)
- [4] Management Michel Crouhy, Dan Galai, Robert Mark：The Essentials of Risk,(2006)
- [5]財政金融委員会調査室 金子隆明：アメリカのサブプライムローン問題 ～アメリカの住宅ローン事情と証券化による金融市場の動揺～,(2007)
- [6]金融情報サイト iFinance
<https://www.ifinance.ne.jp/glossary/finance/fin065.html> (最終閲覧日2019/10/11)
- [7]鹿山昌宏,阿部重夫,武長寛,諸岡康夫：多層ニューラルネットの最適中間層ニューロン数決定法,電気学会論文誌,産業応用部門誌,112(11),1062-1070,(1992)
- [8]枇々木規雄,尾木研三,戸城正浩：小企業向けスコアリングモデルにおける業歴の有効性,(2010)

参考文献

- [9]浅川伸一,“ディープラーニングと中間層の意味”,日本認知心理学会第12回大会書誌, pp28, (2014)
- [10]Michael A. Nielsen: *Neural Networks and Deep Learning*, Determination Press, (2014)
- [11] 田中克明,勝田英紀,萩原統宏:“ディープラーニングと中間層の意味”,日本認知心理学会第12回大会書誌,(2014)
- [12]内田治,ロジスティック回帰分析におけるモデルの適合度指標に関する考察と提案,東京情報大学研究論集,Vol.8 No.1, pp.9-14 (2004)
- [13]天坂格郎,大岩洋之,神尾信,多重共線性の影響と新変数選択法の提案,日本経営工学誌,(1996)
- [14]田中重人,クロス集計表の解釈(2014)
- [15]高橋信(2005),Excelで学ぶコレスポンデンス分析,(オーム社)
- [16]大林準,ロジスティック回帰分析と傾向スコア(propensity score)解析,(2016)

参考文献

[17] Leslie N. Smith, Cyclical Learning Rates for Training Neural Networks(2015)

[18] Lutz Prechelt: *Early Stopping – but when?*, Fakultät für Informatik; Universität Karlsruhe, pp55-69 (1998)

Appendix

APP. 用語について

[I]貸し倒れ。

[II]過去に延滞や破産経験があった場合や、資産や所得と比べた負債水準が高いといった事情により、信用力が低く、好条件では融資が受けられない借り手を対象とした個人向けローン。

[III]デフォルトが発生するなど、貸付け等による投下した資本が返ってこないリスクのこと。

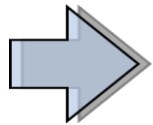
[IV]主要国の金融監督当局で構成するバーゼル銀行委員会が2010年9月に公表した国際的に業務を展開している銀行の健全性を維持するための新たな自己資本規制。

国際的に業務を展開している銀行の自己資本の質と量の見直しが柱で、普通株と内部留保などからなる「中核的自己資本」を、投資や融資などの損失を被る恐れがある「リスク資産」に対して、一定割合以上持つように義務づけるもの。

[V]信用リスクに関する統計モデルのこと。貸し出し先の財務データをもとにデフォルトの判定やデフォルト確率の推定を行う。

APP. 研究背景

- ✓ リーマンショック(2008/9)により世界経済に大きな打撃を与えた。その影響は中南米国にも例外なく与えている。
- ✓ 欧州政府債務危機の影響(2011)により中南米国はGDP成長率が大幅に落ち込み、図12からもわかるように、対ドルレートも2011年を境に大幅に下がっている。



これらのことから中南米国政府は金利政策を行った

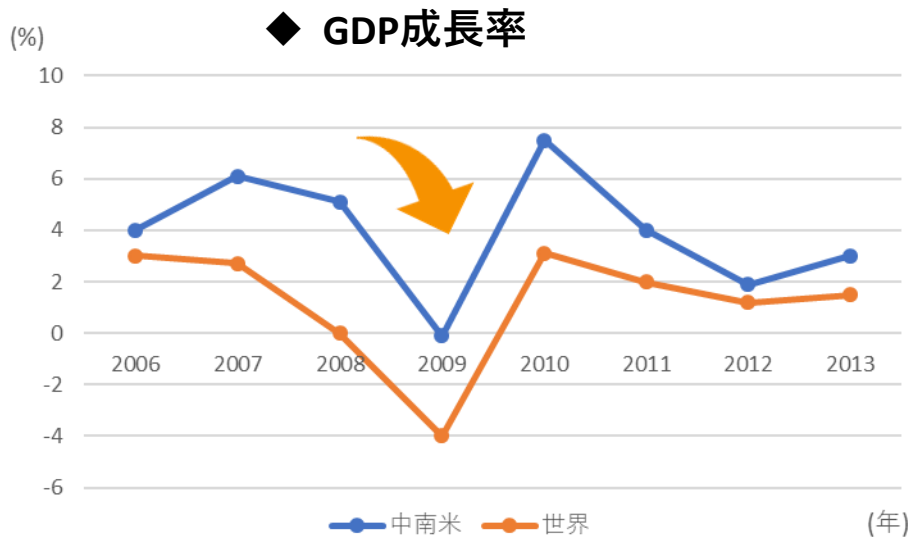


図12 中南米と世界のGDP成長率



図13 レアル通貨の対ドルレート

APP. 研究背景

◆ 就労年数と収入の関係

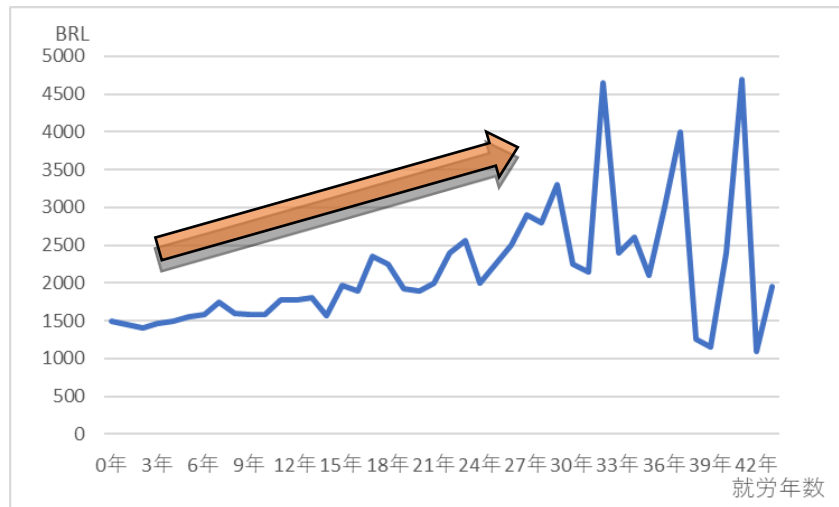


図14 就労年数別平均収入

◆ 就労年数と未払い者率の関係

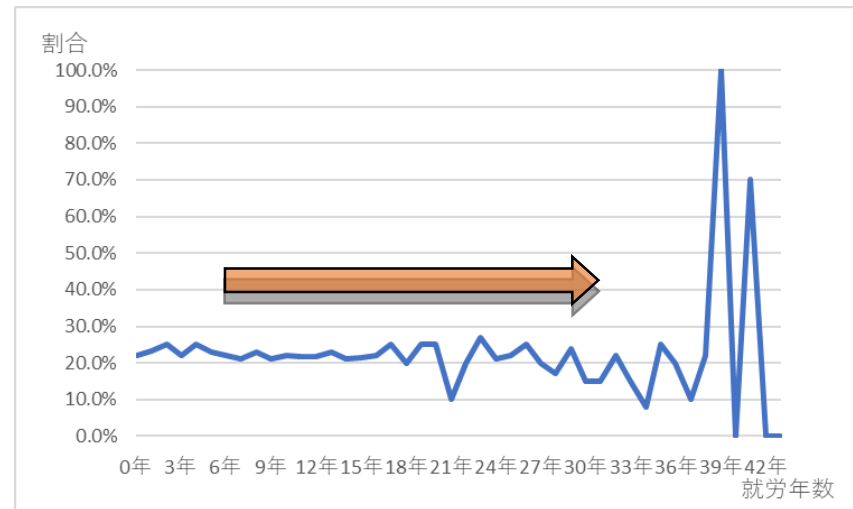


図15 就労年数別未払い者率

収入と未払い者率に大きな関係はないと言える

APP. 研究背景

◆ 学歴別未払い者率

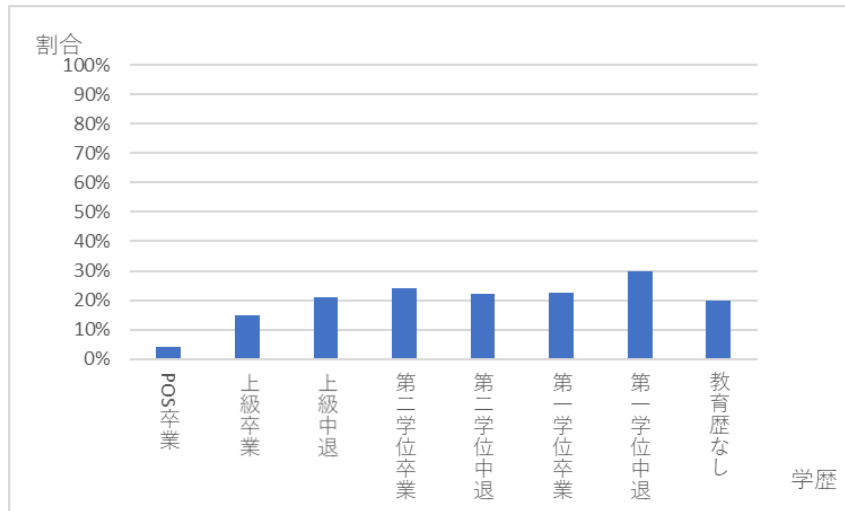


図16 学歴別未払い者率

◆ 職業別未払い者率

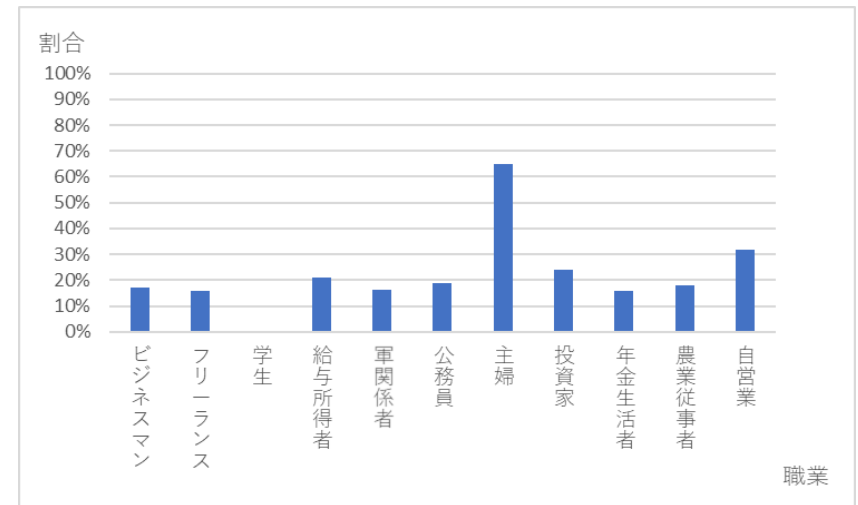


図17 職業別未払い者率

基礎集計だけではどのような属性の人物がデフォルトしやすいのか見分けることは不可能であった

APP. 多重共線性 ~連関係数~

◆ 表6 連関係数結果

	ネガティブ情報	性別	在住
ネガティブ情報	1	0.046	0.089
性別		1	0.068
在住			1

◆ 表7 連関係数結果

	車種
性別	0.264

◆ 表8 連関係数結果

	職業	学歴	家形態	結婚歴
在住	0.593	0.434	0.313	0.194

車種、職業、学歴、家形態、結婚歴の変数を削除した

APP. 多重共線性 ~相関係数~

◆ 表9 相関係数結果

	就労年数	収入	年齢	利子	リスト価格との比	商品価格	頭金	借入金
就労年数	1.000	-0.009	0.100	-0.017	0.008	0.056	0.068	0.032
収入	-0.009	1.000	0.012	0.016	0.000	-0.003	-0.002	-0.004
年齢	0.100	0.012	1.000	0.063	-0.136	0.034	0.013	0.033
利子	-0.017	0.016	0.063	1.000	0.041	0.043	0.059	0.017
リスト価格との比	0.008	0.000	-0.136	0.041	1.000	-0.050	0.109	-0.114
商品価格	0.056	-0.003	0.034	0.043	-0.050	1.000	0.877	0.883
頭金	0.068	-0.002	0.013	0.059	0.109	0.877	1.000	0.556
借入金	0.032	-0.004	0.033	0.017	-0.114	0.883	0.556	1.000
利子金額	0.065	0.003	0.038	0.402	0.165	0.472	0.451	0.392
与信限度額	0.039	-0.001	0.008	-0.003	-0.010	0.209	0.187	0.180
返済能力	-0.002	0.011	-0.011	-0.010	0.095	-0.008	0.010	-0.012
ローン分割数	-0.096	-0.008	0.002	-0.250	-0.225	-0.266	-0.450	-0.040
消費者金融からの照会件数	-0.027	-0.005	-0.021	-0.169	-0.004	-0.046	-0.054	-0.030
ローン会社からの照会件数	-0.069	-0.003	-0.022	-0.123	0.005	0.022	-0.011	0.054

APP. 多重共線性 ~相関係数~

◆ 表10 相関係数結果

	利子金額	与信限度額	返済能力	ローン分割数	消費者金融からの照会件数	ローン会社からの照会件数
就労年数	0.065	0.039	-0.002	-0.096	-0.027	-0.069
収入	0.003	-0.001	0.011	-0.008	-0.005	-0.003
年齢	0.038	0.008	-0.011	0.002	-0.021	-0.022
利子	0.402	-0.003	-0.010	-0.250	-0.169	-0.123
リスト価格との比	0.165	-0.010	0.095	-0.225	-0.004	0.005
商品価格	0.472	0.209	-0.008	-0.266	-0.046	0.022
頭金	0.451	0.187	0.010	-0.450	-0.054	-0.011
借入金	0.392	0.180	-0.012	-0.040	-0.030	0.054
利子金額	1.000	0.127	0.020	-0.350	-0.077	-0.027
与信限度額	0.127	1.000	0.003	-0.099	0.011	0.008
返済能力	0.020	0.003	1.000	-0.027	-0.005	-0.002
ローン分割数	-0.350	-0.099	-0.027	1.000	0.112	0.077
消費者金融からの照会件数	-0.077	0.011	-0.005	0.112	1.000	0.118
ローン会社からの照会件数	-0.027	0.008	-0.002	0.077	0.118	1.000

頭金、借入金、利子金額の変数を削除した

APP. 多重共線性 ~相関比~

◆ 表11 相関比結果

	ネガティブ情報	性別	在住
返済能力	0.069	0.075	0.072
ローン会社からの照会件数	0.123	0.001	0.001
与信限度額	0.013	0.016	0.127
就労年数	0.010	0.005	0.008
消費者金融からの照会件数	0.556	0.007	0.001
ローン分割数	0.064	0.014	0.026
収入	0.131	0.124	0.129
商品価格	0.155	0.145	0.258
リスト価格との比	0.300	0.195	0.389
利子	0.236	0.184	0.186
年齢	0.006	0.005	0.005

消費者金融からの照会件数、商品価格、リスト価格との比、利子の変数を削除した

APP. 分析補足

◆ 表12 多項ロジスティック回帰結果

	p値
ネガティブ情報	0.00
性別	0.00
在住	0.00
就労年数	0.00
分割数	0.00
ローン会社からの照会件数	0.00

◆ 表13 中間層による精度比較

	1つ	2つ	3つ
精度	0.78	0.78	0.78

◆ 表14 ミニバッチサイズによる精度比較

	64	1024
精度	0.78	0.78

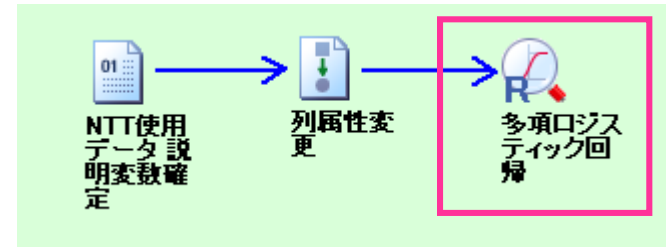


図18 多項ロジスティック回帰フロー図

多重共線性、P値を考慮した6変数に関してVRPの「**多項ロジスティック回帰**」でP値を出したところ全て0.00であった。

1~3つの中間層を検証したところ精度に差は出なかった。

ミニバッチサイズについて64と1024を検証したが精度に差はでなかった。