

審査AIモデルのご紹介

～審査AIモデルによる新たな審査スタイル～

2021年1月

人工知能
(AI : artificial intelligence)

－使い方の定義－

機械学習をはじめとする**何らかの入出力システム**に、
人間的な仕事をさせると、人工知能と呼ばれることが多い※

機械学習
(machine learning)

－作り方の定義－

従来のモデルの主流

世界チャンピオンに圧勝した
囲碁AI「Alpha Go」は、碁
盤を画像のように認識する
CNNを採用

ニューラルネットワーク
(NN : neural network)

ディープラーニング
(deep learning)

多層ニューラルネットワーク (DNN)
再帰型ニューラルネットワーク (RNN)
畳み込みニューラルネットワーク (CNN)

<近年急速に進歩した領域>

サポートベクターマシン

ランダムフォレスト

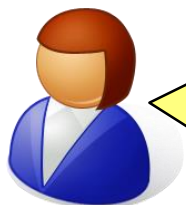
ロジスティック回帰

<以前から存在した領域>

名人に勝利した将棋AI「ポナンザ」(旧バージョン)は、
ロジスティック回帰であり、
ディープラーニングではない

審査や格付のために使う統計モデルも立派な「AI」ですから、銀行は20年前からAIの使い手なのです。

(※2017.12.15 ファイナンスリスクのモデリングと制御IV (統計数理研究所 山下教授) による)



法人融資の分野に関連するAI活用の記事の多くは、ようやくサービス開始を告げるものも見られるようになってきましたが、なかにはすでに撤退を決めたところもあり、従来の融資と大きな違いを示すような内容は多く見つけられません。

【事業性融資にかかる金融機関のAI活用リリース（2020年6月まで）】（出所：Webサイト等より弊社調べ（7/1））

段階	銀行	関連企業名	タイトル
サービス開始	三菱UFJ銀行	-	AI融資で金利優遇 新型コロナ対応（2020年3月18日）
サービス開始	みずほ銀行	-	中小企業向けの新しいレンディングビジネスへの取り組み みずほスマートビジネスローンの取り扱い開始（2019年4月16日）
サービス開始	りそな銀行	-	AIが信用力審査 口座データ基に中小向け融資（2019年12月25日）
サービス開始	-	損保ジャパン	企業の資金調達、保険で支援 AIにて融資可能額算出（2020年6月24日）
検証	愛知銀行	エメラダ	人工知能で倒産リスクを判定する融資先の与信分析ツールを共同検証（2020年6月4日）
サービス終了	-	マネーフォワード	マネーフォワード、融資審査にAI活用（2018年7月11日） ⇒ 2020年8月にサービス終了（2020年6月8日）



事業性融資（法人融資）の分野は、個人にくらべてデータの数が圧倒的に少ないので、精度の高いAIの開発が格段に難しいものとされています



20年前から使っている統計モデルによるスコアリング審査を、あらためて「AI」と名付けてリリースしているケースもありそう

<p>現状認識</p>	<ul style="list-style-type: none"> ✓ 近年の機械学習は、「非構造化データ」※の解析に威力を発揮しており、具体的には、画像解析、言語解析、音声解析、動画解析等の分野にてブレイクスルーが見られる ✓ 近年のハードウェアの技術革新により、「非構造化データ」を、分析可能な「構造化データ」に変換することが可能になった <p>※ テキストデータ、画像データ、音声・動画データなど数値データ以外の全てを指す</p>
<p>課題</p>	<ul style="list-style-type: none"> ✓ 財務データに対する分析結果からも明らかなように、すでに一定の分析成果が上がっている構造化データにおいては、<u>手法だけを機械学習に置き換えても、追加的に得られる成果は限定的</u> ⇒ <u>データが同じなら、手法を変えるだけでは高い成果は得られない</u> ✓ 「機械学習革命」の本質は、膨大な量がありながら、従来の技術では分析しきれなかった「非構造化データ」を新たに分析対象に加えられるという、「データソース革命」にある ⇒ <u>新たなデータソースにアクセスして初めて、高い成果が期待できる</u> (例) コールセンターの音声、担保物件の写真、稟議書意見欄

【参考：機械学習によるモデルの性能（AR）】

	構築用データ	評価用データ
従来の統計モデル	0.65748	0.72356
ニューラルネットワーク	0.74046	0.73413
ランダムフォレスト	0.81284	0.74355

(弊社調べ)

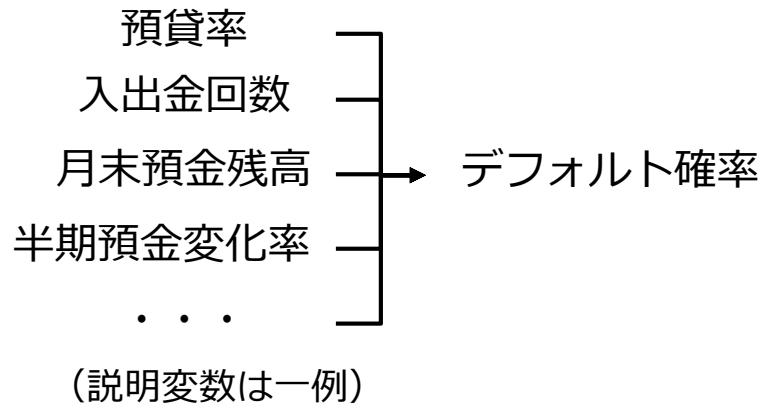
最新の機械学習手法は、使えるデータが、これまで使われていなかった「非構造化データ」に広がったのが特徴です。活用するデータが変わらなければ、得られる成果も今までと変わりません。



(ARは0と1の間をとる統計量。1に近いほど高性能)

新旧AIモデルの比較（回帰モデルとニューラルネットワーク）

【旧AIの例】ロジスティック回帰モデル



<メリット>

- 説明変数（口座情報による指標）と被説明変数（デフォルト確率）との関係がわかりやすい
- データ数量に見合った、相応の性能と頑健性の確保が可能
- システム実装が比較的容易

<デメリット>

- モデルの性能には一定の限度
- ヒトによる説明変数の洗い出しを要する（データに対する経験的なノウハウの蓄積が十分でないデータには向かない（cf.財務指標））
- 指標の複雑化、指標数の拡大によって性能は上がる可能性があるが、同時に説明可能性は低下する
- 線形モデルの場合、残高項目の寄与度が極端に高いため、それ以外の情報は説明変数に採用しづらい

【新AIの例】畳み込みニューラルネットワーク（CNN） （HALCA-A（CNN版））

縦×横の2次元データ（非構造データ）として処理

	201501	201502	201503	201504	201505	201506	201507	201508	201509	201510	201511	201512	201601
各種 口座情報 項目	xx,xx	xx,xx	xx,xx	xx,xx	xx,xx	xx,xx	xx,xx	xx,xx	xx,xx	xx,xx	xx,xx	xx,xx	xx,xx
	xx,xx	xx,xx	xx,xx	xx,xx	xx,xx	xx,xx	xx,xx	xx,xx	xx,xx	xx,xx	xx,xx	xx,xx	xx,xx
	xx,xx	xx,xx	xx,xx	xx,xx	xx,xx	xx,xx	xx,xx	xx,xx	xx,xx	xx,xx	xx,xx	xx,xx	xx,xx
	xx,xx	xx,xx	xx,xx	xx,xx	xx,xx	xx,xx	xx,xx	xx,xx	xx,xx	xx,xx	xx,xx	xx,xx	xx,xx
	xx,xx	xx,xx	xx,xx	xx,xx	xx,xx	xx,xx	xx,xx	xx,xx	xx,xx	xx,xx	xx,xx	xx,xx	xx,xx
	xx,xx	xx,xx	xx,xx	xx,xx	xx,xx	xx,xx	xx,xx	xx,xx	xx,xx	xx,xx	xx,xx	xx,xx	xx,xx
	xx,xx	xx,xx	xx,xx	xx,xx	xx,xx	xx,xx	xx,xx	xx,xx	xx,xx	xx,xx	xx,xx	xx,xx	xx,xx
	xx,xx	xx,xx	xx,xx	xx,xx	xx,xx	xx,xx	xx,xx	xx,xx	xx,xx	xx,xx	xx,xx	xx,xx	xx,xx
	xx,xx	xx,xx	xx,xx	xx,xx	xx,xx	xx,xx	xx,xx	xx,xx	xx,xx	xx,xx	xx,xx	xx,xx	xx,xx
	xx,xx	xx,xx	xx,xx	xx,xx	xx,xx	xx,xx	xx,xx	xx,xx	xx,xx	xx,xx	xx,xx	xx,xx	xx,xx
	xx,xx	xx,xx	xx,xx	xx,xx	xx,xx	xx,xx	xx,xx	xx,xx	xx,xx	xx,xx	xx,xx	xx,xx	xx,xx
	xx,xx	xx,xx	xx,xx	xx,xx	xx,xx	xx,xx	xx,xx	xx,xx	xx,xx	xx,xx	xx,xx	xx,xx	xx,xx
	xx,xx	xx,xx	xx,xx	xx,xx	xx,xx	xx,xx	xx,xx	xx,xx	xx,xx	xx,xx	xx,xx	xx,xx	xx,xx
	xx,xx	xx,xx	xx,xx	xx,xx	xx,xx	xx,xx	xx,xx	xx,xx	xx,xx	xx,xx	xx,xx	xx,xx	xx,xx
	xx,xx	xx,xx	xx,xx	xx,xx	xx,xx	xx,xx	xx,xx	xx,xx	xx,xx	xx,xx	xx,xx	xx,xx	xx,xx
	xx,xx	xx,xx	xx,xx	xx,xx	xx,xx	xx,xx	xx,xx	xx,xx	xx,xx	xx,xx	xx,xx	xx,xx	xx,xx
	xx,xx	xx,xx	xx,xx	xx,xx	xx,xx	xx,xx	xx,xx	xx,xx	xx,xx	xx,xx	xx,xx	xx,xx	xx,xx
	xx,xx	xx,xx	xx,xx	xx,xx	xx,xx	xx,xx	xx,xx	xx,xx	xx,xx	xx,xx	xx,xx	xx,xx	xx,xx
	xx,xx	xx,xx	xx,xx	xx,xx	xx,xx	xx,xx	xx,xx	xx,xx	xx,xx	xx,xx	xx,xx	xx,xx	xx,xx

<メリット>

- モデルの性能は回帰モデルよりも改善
- ヒトによる説明変数の洗い出しは不要
- 長期間、高頻度の時系列情報も入力データとして採用しやすい

<デメリット>

- 説明変数と被説明変数との因果関係を把握することは難しい（ブラックボックス化）
- システム実装の難易度は比較的高い



CNNの手法によるHALCA-Aでは、口座情報のモデル化で、高い性能を実現しています

審査AIモデルとは、統計的手法によらない機械学習による手法で構築した審査モデル（＝スコアリングモデル）です。



各種情報を統合して構築した審査モデル。複数のHALCAモデルを組み合わせる。

審査AIモデル
HALCA(ハルカ)

HALCA-A
(Account)
口座情報モデル

動態情報（預貸、商流含む）を用いたモデル



HALCA-G
(Goodwill)
テキストモデル

交渉記録、稟議書、アバターが収集する会話の情報等を用いたモデル



HALCA-F
(Finance)
長期財務モデル

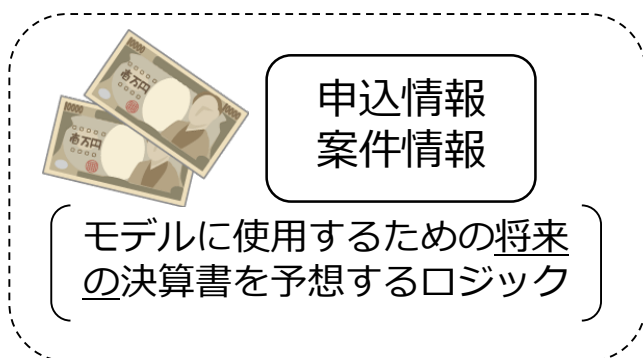
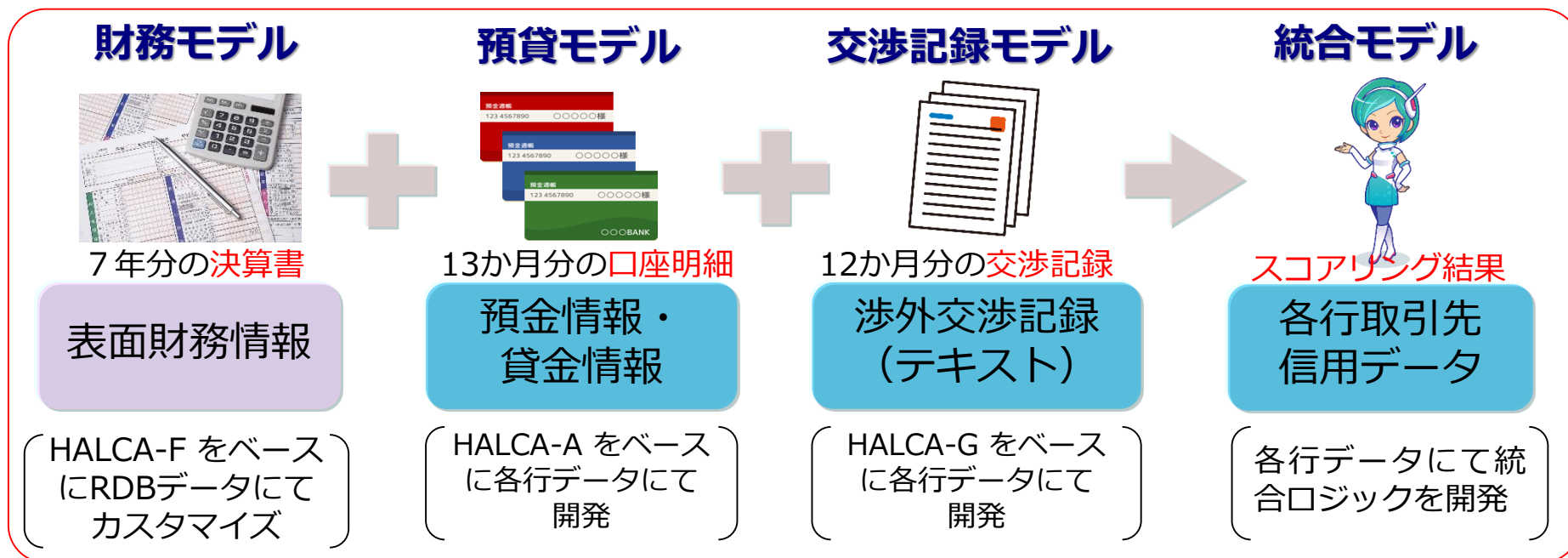
長期間の財務情報を用いたモデル



以降において、AIモデルとは「統計的手法によらない機械学習による手法で構築したモデル」を指します。

非構造化データによる審査AIモデルの事例

3つの異なる情報を用いたAIモデルを統合して貸出先の最終的な評価を決定する「統合モデル」を構築します（交渉記録に加え、財務データ、預貸データも非構造化データとして処理します）。



<凡例>
モデルの開発に使用するデータ

- 御行 ... 各行のデータ
- RDB ... RDBのデータ
- ... 参照のみ

審査AIモデルによる評価と信用格付（人間の評価）を組み合わせ、審査手続きの合理化・省力化と、貸出採算向上を目指す取り組みが始まっています

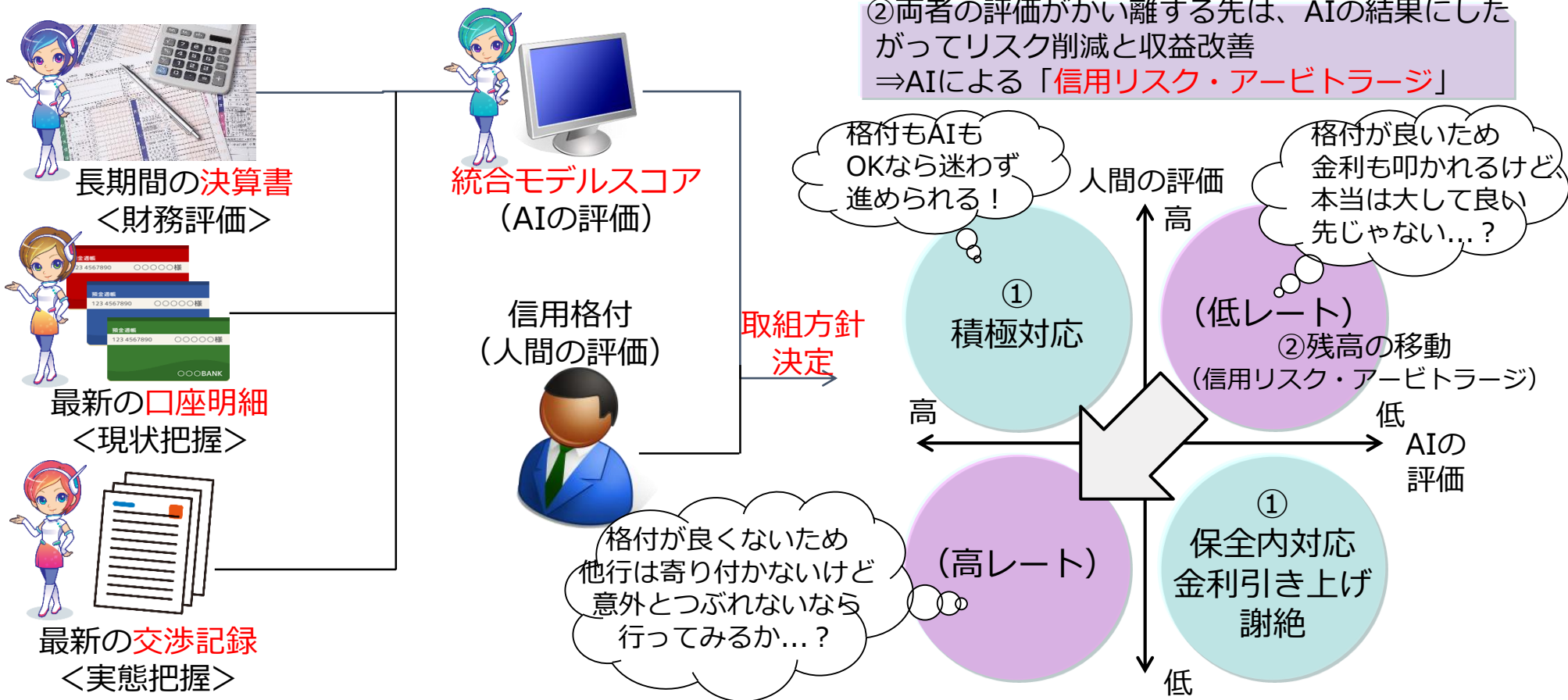
取り組み方針の判定

【審査AIモデルを活用した案件審査の流れ】

統合モデルによるスコア（AIによる評価）と、従来の信用格付（人間による評価）の両者を比較することで、相互に牽制を働かせながら、効率的・合理的な貸出先の審査を実現します

①両者の評価がそろった先は、審査手続きを簡略化
⇒AIによる審査の**合理化・省力化**

②両者の評価がかい離する先は、AIの結果にしたがってリスク削減と収益改善
⇒AIによる「**信用リスク・アービトラージ**」



【補足】AIモデルはブラックボックスか？

- ❑ 衣料品卸売業のデフォルト先の事例。弊社が提供する統計モデル（中小企業クレジット・モデル 2008）の評価は、10段階で上から3番目に対して、HALCA-Fの評価は10段階の最下位
- ❑ 自己資本比率50%超、現預金も相応に保有しており、従来のスコアリングモデルでは低い評価になりにくい、ヒトが見れば「何かがおかしい」と気付くヒントのある財務
- ❑ HALCA-Fの頭の中は「ブラックボックス」でも、結果は「見ればわかる」

【デフォルトサンプルの例（単位：百万円）】

貸借対照表（借方）

	直近期	前期	前々期
現金・預金	8,063	6,730	8,060
受取手形	1,179	1,205	1,052
売掛金	5,694	5,797	6,750
当座資産計	14,936	13,733	15,862
製商品、仕掛品	3,071	4,967	5,981
棚卸資産計	3,104	5,010	6,034
短期貸付金	39	44	298
未収入金	124	138	14
その他流動資産	9	6	75
貸倒引当金_流動資産	-48	-48	-81
流動資産合計	18,165	18,896	22,229
有形固定資産計	191	191	202
固定資産合計	13,936	13,463	11,061
繰延資産合計	80	175	321
資産合計	32,181	32,534	33,611

※投資等に不良資産を計上???

貸借対照表（貸方）

	直近期	前期	前々期
支払手形	227	215	527
買掛金	631	772	1,099
短期借入金	6,500	6,850	8,000
未払金	378	358	624
賞与引当金等	147	132	193
その他流動負債	241	191	129
流動負債合計	8,298	8,519	10,576
長期借入金	3,620	930	0
固定負債合計	5,620	5,930	5,000
負債合計	13,918	14,449	15,576
資本金	56	56	56
資本合計	18,263	18,085	18,035
割引手形	0	0	0



粉飾チェックのツールとしても、銀行内部で使われています

損益計算書

	直近期	前期	前々期
売上高	27,378	27,984	33,301
売上原価	21,457	22,552	26,260
売上総利益	5,920	5,432	7,041
販売費・一般管理費	4,923	5,408	6,289
営業利益	998	24	752
営業外収益	518	655	390
支払利息・割引料	246	261	222
営業外費用	956	481	504
経常利益	560	198	638
税引前当期利益	522	284	663
法人税等充当額	232	123	262
当期利益	291	162	401
次期繰越利益	9,939	9,761	9,711
当期減価償却費	32	36	43
従業員数	140	140	140
有利子負債	12,120	12,780	13,000

※本業以外で最終利益が大きく変動