

【統一論題報告・論文】

未来のディスクロージャーとその分析

坂上 学 (法政大学)

企業のパブリック・ディスクロージャーは、定性的な情報だけでなく、非定型化・非構造化した情報の開示が拡大しており、この傾向は将来も続くであろう。過去においては、もっぱら財務諸表データを中心とした単純な分析手法だった時代もあったが、近年ではファイナンス理論や計量経済学を駆使した企業評価へと変貌を遂げている。今後は非定型・非構造化データの情報開示が今後も増大していくことを鑑みると、新たな分析手法へと変容していくことになるだろう。このような中で、データマイニングや機械学習などの手法を経営分析に応用しようという動きが散見されるようになってきている。そこで本稿では、近年発展著しいディープラーニングの手法について、その基本的な仕組みを解説し、ビッグデータへの対応もにらみながら、経営分析への応用可能性について検討することにする。

キーワード：非定型データ、非構造化データ、データマイニング、機械学習、人工知能（AI）、ディープラーニング、ニューラルネットワーク、ビッグデータ

I. はじめに

本稿の目的は、これまでのディスクロージャーの技術的な展開と、近い将来における技術革新がもたらす経営分析手法の影響について、多少の推測を交えながら論じることである。簡単に言えば、非定型データあるいは非構造化データの増大と、それに対応した新たな経営分析手法の必要性により、経営分析の主たる方法が大きく変わる可能性を指摘する。

現在の経営分析は、従来の会計情報を中心とする財務諸表分析から、ファイナンス理論を導入し資本市場との関係に注目する企業評価へと移行している。しかしながら現代においては、

会計情報や金融情報のみならず、インターネット等を通じて、フォーマットにとらわれない自由文書や図表など様々な非定型データないしは非構造化データを入手することができる。

非定型データないしは非構造化データを分析するため、1990年代よりデータマイニングやテキストマイニングなどの技術が登場し、少なからず会計領域にも影響を与えてきた(坂上 2015)。データマイニングには、統計解析の手法が用いられるが、それに加え機械学習といった人工知能の手法も使われる。機械学習といっても様々なものがあるが、その代表的なものに、ニューラルネットワーク、ベイジアンネットワーク、遺伝的アルゴリズム、サポート・ベクター・マシンなどがある。

Ⅱ. 先行研究

1. データマイニング研究の展開

データマイニング手法を用いた最初期の研究としてはFanning, Cogger & Srivastava (1995)を挙げることができる。経営者による不正発見の問題について、Loebbecke, Eining & Willingham (1989)は不正発見のための概念モデルを応用したロジック・アプローチを採用しているが、この方法に比べニューラルネットワークの手法を応用した方が、不正発見の可能性が高まることを示した。

Fanning & Cogger (1998)はさらに、ニューラルネットワークの手法を、公表された財務諸表データに適用し、不正発見の助けとなることを明らかにした。

Kotsiantis, Koumanakos, Tzelepis & Tampakas (2006)は、ベイジアンネットワークの手法を初めて導入したもので、2000-2001年のギリシャ企業の財務諸表を対象として不正発見（実際には倒産企業と非倒産企業の判別）をおこなっている。

Hoogs, Kiehl, Lacombe & Senturk (2007)は、遺伝アルゴリズムを用いた研究で、財務諸表の不正を発見するために企業の財務行動を特徴付ける異常スコアというものを定義し、このスコアを用いて遺伝アルゴリズムにより発見させようと試みた研究、Ravisankar, Ravi, Rao & Bose (2011)は、サポート・ベクター・マシンを取り入れた研究である。

これらの研究に共通するのは、企業の不正会計の発覚のサロゲートとして、倒産の発生という事象を用いていることと、Altman (1968)のZ-scoreに代表される倒産発生の判別分析をより精緻なものにしようとして、データマイニングの手法を応用した点である。これらはすべて、人工知能 (Artificial Intelligence : AI) の領域

における機械学習のアルゴリズムであり、より人間に近い判断ができる分類器 (classifier) として機能するように考えられてきたものだ。

2. 機械学習の深化 (ディープラーニング)

近年の人工知能における機械学習の精度が増したことにより、ディープラーニングがここ数年大きな注目を浴びている。ディープラーニングについては、結論から言えば、アルゴリズムについてはニューラルネットワークについての知識、そしてそれを扱うためのコンピュータ・システムについての知識が必要となる。

Ⅲ. ディスクロージャーの技術的変遷

会計ディスクロージャーの初期段階では、紙媒体による財務諸表の提供が主たる開示方法であった。コンピュータの普及とネットワーク環境が整うことをうけて、米国SECは開示書類の電子化を推進し、1996年よりすべての開示書類がインターネットを通じて入手できるようになった。

しかしながら、単に印刷媒体で提供されていたデータを電子化しただけでは、データの2次利用に大きな問題があることが判明した。それを克服するためにXBRLが登場し、企業の財務データの定型化・構造化が一気に進むことになった (坂上 2012)。

財務データの定型化・構造化は進んだものの、企業を評価する上で財務情報の役割はそのウエイトを下げてきたといわれる。その一方で、近年の統合報告という動向に見られるように、いわゆる環境・社会・ガバナンス (ESG) 情報など多様な情報が同時に提供されるようになった。この結果、非定型・非構造化データが増大することになった。環境報告書やCSR報告書といった情報には、文章による記述や図表が多く含まれている (図表1)。これらの非定型・非

図表1 環境報告書の記述例 (法政大学環境センター 2016, p. 3)



構造化データを理解し判断するためには、やはり人間による作業が必要とされていた。

IV. 人工知能を利用した経営分析技術の展開

非定型データないしは非構造化データについて、とりわけ文書データについては、データマイニングの手法の応用でもあるテキストマイニングの手法の応用がこれまで適用されてきた。有価証券報告書に含まれるテキスト情報を、形態素解析をすることによってすべての語句をリストアップし、頻度分析を行った結果と財務データとを突き合わせることで、企業の倒産予知を行うといった方法(白田・坂上 2008)や、アンケート調査結果から何が鍵となる概念であるかについて、TF/IDFを用いて析出する方法(坂上・清水・島本 2007)などがある。

ディープラーニングの手法が登場したことで、近年の人工知能における機械学習が深化し、いわゆるディープラーニングがデータ分析に応用されることによって(斎藤 2016, 橋本 2017)、新たな段階に入りつつあることが認識されるようになってきた。一部では人間をも凌駕する性能も見せるようになり、人工知能の将棋や囲碁の力量が、もはやプロの棋士をも凌駕するのは時間の問題と言及されるようになってきている。

人工知能を使えば、現在のアナリストよりもまともな分析レポートが出せるはずだという予測もしばしば耳目を集めるようになってきている。富士通研究所(2012)よりビッグデータを利用して分析シナリオを自動生成するシステムの開発がアナウンスされた。これらの近年の動向は経営分析手法の将来に大きなインパクトを与えざるはずだ。

ディープラーニングの会計研究への応用については、坂上 (2017) がその可能性について議論をおこなっている。ここではそこでの議論を参照しつつ、以下において齊藤 (2016) および橋本 (2017) の説明を参照しながらディープラーニングの基礎を確認することにしよう。

V. 機械学習・ディープラーニングの基礎

1. ディープラーニングはブラックボックスではない

前述した内容に関連するニュースや雑誌記事などでは、「ビッグデータ」、「人工知能」、「機械学習」、「ディープラーニング」といった言葉ばかり先行しているきらいがある。実際にその内容についての理解は経営分析領域では意外と進んでいない。また「ディープラーニングは、中身はブラックボックスだから」といった言説も多々見受けられる。このような状況は、メインフレーム・コンピュータ上で動作する統計パッケージが登場した時とよく似ている。当時の統計学者の多くが、コンピュータの動作原理やプログラミングの知識に欠けていたため、コンピュータ上で動作する統計パッケージはブラックボックスであり、それゆえ計算結果につい

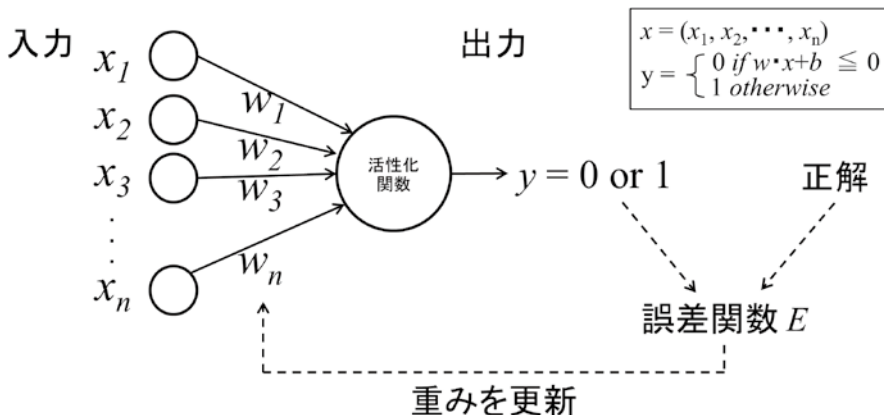
て懐疑的な目が向けられていた状況であったことによく似ている。

しかしながら、現在の代表委的なディープラーニングの手法である、ニューラルネットワークと特徴表現抽出の手法において、アルゴリズムと誤差関数がしっかりと定義されており、でたらめな動作、あるいはわけのわからない動作をしているということでは決してない。ただそれを理解しようとしな (できない) 人たちが「ブラックボックス」という表現でごまかしているに過ぎない¹⁾。

実証研究のように、計量経済学的手法を用いて統計解析をおこなう場合、程度の差はあれ統計学の知識を習得しているはずであり、また回帰分析の結果得られた偏回帰係数の統計的有意性の意味についても理解しているはずである。2 値変数が使われているならばロジスティック回帰分析(ロジット分析とも)を使ったり、プロビット回帰分析を使ったりするといった知識も持っているはずだ (そして、この知識はディープラーニングの理解にも大いに役立つことになる)。

実証研究をおこなう者は、計量経済学や統計学の知識を学び、それに加えてツールとしてStataやRのような統計ソフトを利用するよう

図表2 単純パーセプトロン



※橋本 (2017, 204-205) の図に齊藤 (2012, 42-43) の説明を加味したもの

に、ビッグデータと機械学習を応用したアナリティクスとよばれる分析手法を応用しようとするならば、ディープラーニングなどの機械学習に関する知識を学び、それに加えてツールとしてChainerなどのディープラーニング・フレームワークを利用することが求められる。AIの専門家になる必要は決していないが、少なくともディープラーニングの基本的な知識を習得しておき、内部でどのような動作をしているかを知ったうえで利用することが肝要であろう (坂上 2017)。

2. 機械学習の展開

まずは初期の形式ニューロンの話から始めることにしよう。McCulloch & Pitts (1943) は、形式ニューロン (人工ニューロン) を初めて提唱したが、これは神経細胞の振る舞いにヒントに論理計算をおこなうモデルを示したものである。その後、Rosenblatt (1958) は、形式ニューロンのアイデアをもとにパーセプトロン (perceptron) というモデルを開発した。これは、入力層と出力層の2層からなり、「単純パーセプトロン」と呼ばれるモデルである。

しかしながらこのモデルは、Minsky & Papert (1969) により、線形非分離問題を解けないことが指摘され、一時期下火となってしまった。線形非分離問題を一言で説明することは難しいが、論理演算においてXOR演算 (排他的論理和) の結果を一次線形式を用いて分離することは不可能で、2次式 (以上) の曲線でしか分離できないといった例で端的に示される。

やや教科書的な記述になるが、XOR演算について、それが何を意味するかについて簡単に説明をおこなうことにしよう²⁾。論理演算の基本は、AND演算 (論理積)、OR演算 (論理和)、NOT演算 (否定) の3つであり、さらにこれら3つの演算を組み合わせるとNAND演算 (否定論理積)、NOR演算 (否定論理和)、XOR演算 (排

他的論理和) の3つがある。

AND演算は、ともに真理値が1の場合に1を出力する演算である。OR演算はどちらかの真理値が1の場合に1を出力する演算である。NOT演算は、真理値が1であれば0を、真理値が0であれば1を出力する演算である。NAND演算はAND演算の否定なので、ともに真理値が1である場合以外のすべてにおいて1を出力する演算である。NORはOR演算の否定なので、どちらかの真理値が1である場合以外のすべてにおいて1を出力する演算である。やや難しく感じられると思われるのはXOR演算で、異なる真理値を持つ場合のみに1を出力する演算である。

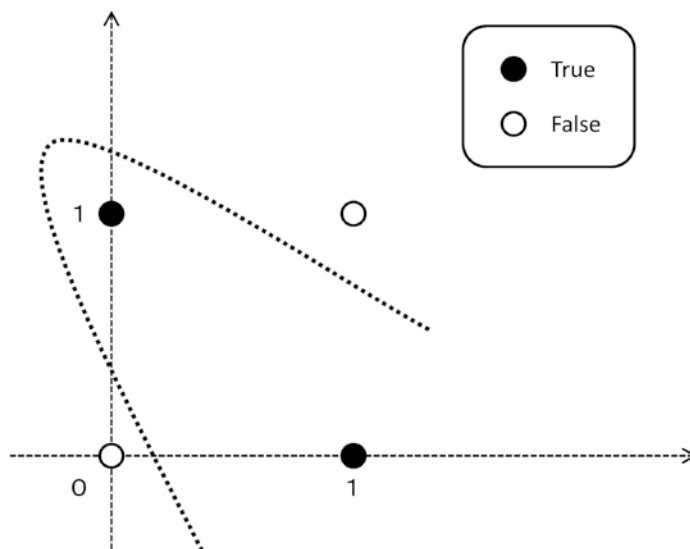
このうちXOR演算の結果は、線形非分離問題が発生する典型的な例となっている。論理的な判別を下すためには上記のような論理演算をおこなえることが条件となるが、そのためにはXOR演算ができないのは致命的である。これを分かりやすく図示したものが図表3である。そして、この限界を克服するために、Rumelhart, Hinton & Williams (1986) の誤差逆伝播学習法を導入した多層パーセプトロン (Rumelhart & McClelland 1986) が提唱されることになった。

3. 誤差逆伝播学習法とシグモイド関数

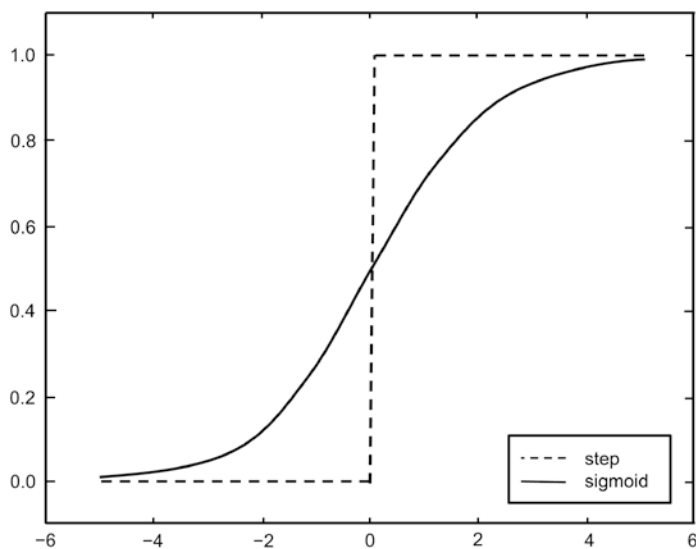
単純パーセプトロンでは0または1の2値をとればよいので、活性化関数はステップ関数でよかった。しかしながら誤差逆伝播学習法では、誤差関数をパラメータで微分することによって、パラメータの修正方向を決め、パラメータの修正を繰り返すという勾配法を拡張したものであるため、多層パーセプトロンで用いられる活性化関数では、勾配計算をするために微分できることが必要となる。このため、シグモイド関数が導入される。

西尾 (2015, p. 30) によれば、単純パーセプトロンの進化の方向大きく2つの流れがある。1つは「能力拡張路線」で、ロジスティック回

図表3 XORは線形で分離することはできない



図表4 ステップ関数とシグモイド関数の違い



帰, サポート・ベクター・マシン, Confidence-Weighted Linear Classifierへと続くものである。もう1つは, 「積み重ね路線」で, 多項ロジスティック回帰, ニューラルネットワーク, ディープラーニングへと続くものである。

能力拡張路線の代表であるロジスティック回帰 (logistic regression) は, 単純パーセプトロ

ンのように2値分類するのではなく, 確率値を出力する機械学習アルゴリズムである (橋本2017, p. 206-208)。出力値は, 0から1の間の確率値に写像されるが, この時にシグモイド関数が用いられる。これに対し, 積み重ね路線の代表である多項ロジスティック回帰 (multivariable logistic regression) は, 出力yが多次元となる

ロジスティック回帰で、ただ単にロジスティック回帰を次元数の数だけ並列に実行する機械学習アルゴリズムである (橋本 2017, p. 208-209)。誤差関数 E は、いずれも対数尤度 $L(w)$ を用いて定義される³⁾。

4. ニューラルネットワークへの展開

ニューラルネットワークは、いくつも単純パーセプトロンを組み合わせたものである (橋本 2017, p. 211)。代表的なものは階層型ニューラルネットワークで、入力層、中間層、出力層という3つのユニットから構成される (図表5)。入力層は、入力された値を中間層の各ユニットに伝搬するだけの動作をおこない、中間層と出力層のユニットは、複数のユニットから出力された値を入力として、ある関数により値を計算し出力をする。複数の入力値 (x_1, x_2, \dots, x_n) から総入力値 u を計算し、その u を関数 f に適用して出力値 $z (=f(u))$ を計算する (橋本 2017, p. 213-214)。

ニューラルネットワークの学習は、誤差関数 E (損失関数とも) を指標として、誤差関数 E の値が小さくなるように、重みパラメータを更新する。重みパラメータの更新の際に、重みパラメータの勾配を利用し、勾配方向に重みの値を更新する作業を繰り返しおこなう。この重みパラメータの勾配は、数値微分によって求めることができるが、計算に多くの時間を必要とす

る。そこで勾配を高速に求めることができるように登場したのが、前述した誤差逆伝播学習法である。

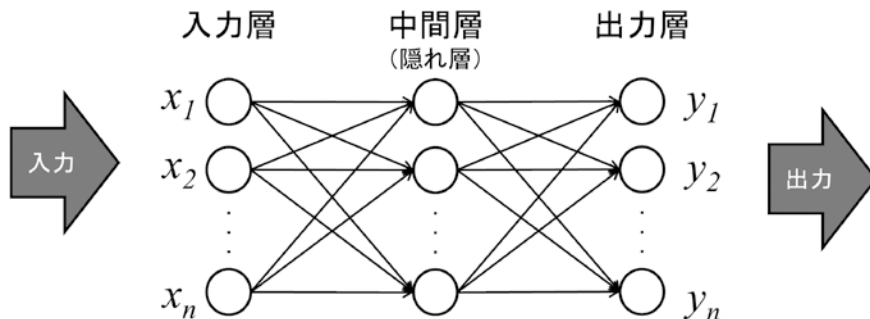
5. ディープラーニングへの展開

深層学習 (deep learning) といわれる機械学習アルゴリズムの1つとして有名なのが、畳み込みニューラルネットワーク (convolutional neural network) である。これは、畳み込み層 (convolutional layer) とプーリング層 (pooling layer) の組み合わせを何度か繰り返すことで、特徴的な要素を抽出するというものである⁴⁾。最終段では多層パーセプトロンに受け継がれ、何らかの値を出力する (図表6)。

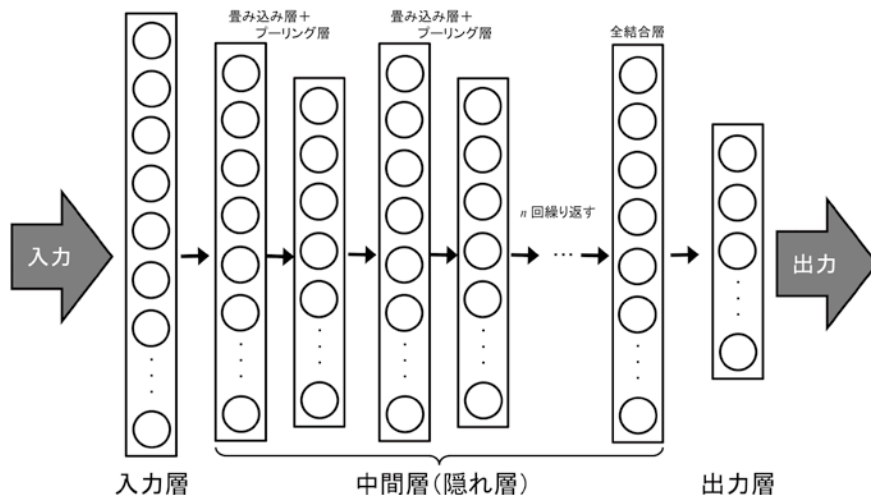
VI. おわりに - 経営分析からビジネス・アナリティクスへ

以上、経営分析領域へのAIの導入の歴史について概観し、ディープラーニングの基礎としてニューラルネットワークから最近のディープラーニングに至る展開を駆け足に説明したが、今後はビッグデータとディープラーニングといった手法をいかに使いこなすかということが経営分析領域の研究に求められるようになるだろう。監査実務の世界では既にその兆候がみられており、主要な監査法人はどれもビッグデータと監査データアナリティクスといったワードを用いながら、今後の監査のあり方が語るよう

図表5 ニューラルネットワーク



図表6 畳み込みニューラルネットワーク



になっている⁵⁾。

ところでビッグデータとは、どのようなものであろうか。Wikipediaの記述によれば「市販されているデータベース管理ツールや従来のデータ処理アプリケーションで処理することが困難なほど巨大で複雑なデータ集合の集積物を表す用語」ということである。物理学の世界では、しばしばペタバイト（ゼロが15個、1,000テラバイト）級のデータを扱うことがあるが、これは会計学を含む社会科学領域の学問では未知の領域と言えるだろう。普通の実証研究などで用いる財務データ（たとえば日経NEEDS）は、1ギガバイト（ゼロが9つ）に満たないことが多く、データサイエンス領域で扱うデータの100万分の1程度にすぎない。10GBを超えるデータになると、一般的なパソコン上で動作するツール（パソコン用データベースソフトや表計算ソフト）では著しく扱いづらく、1TBを超えるデータの場合、もはやパソコンで扱うことは不可能となる。

一般的な文系研究者が扱えるデータは、せいぜい100GBまでというのが正直な印象である。ビッグデータとディープラーニングといった領域に踏み出すには、コンピュータもパソコン・

レベルではなく、ワークステーション・レベルのものが最低限必要となるし、扱うOSもWindowsではなくLinuxやFreeBSDなどのUnixライクOSの利用も視野に入れる必要が出てくる。

したがって、いきなりディープラーニングに挑むというよりも、データマイニングの単純なアルゴリズム（ニューラルネットワーク）から始めるのが良いのではないかというのが筆者の偽らざる気持ちである。テキストマイニングをするにしても、まずは形態素解析と頻度分析（TF/IDF等）あたりから始めるのが無難かもしれない。しかしながら、これらの基本的な分析手法の積み重ねの上に、ディープラーニングの世界が広がっているのである。

経営分析（Business Analysis）と呼ばれていた財務諸表分析を中心とする従来の伝統的分析手法は、ファイナンス理論と資本市場のデータを駆使した企業評価（Business Valuation）と呼ばれる手法へと移行し今日に至っている。今後は、ビッグデータとディープラーニングを用いたビジネス・アナリティクス（Business Analytics）への移行が進むと予想される。

現在、有価証券報告書のデータはすべてXBRLによってタグ付けされた状態で開示されており、

XBRLの知識さえあれば財務数値だけでなく、文字情報もすべて特定の記述を抽出し、テキストマイニングをかけるといったことができるようになってきている。全上場企業の有価証券報告書のXBRLデータを1年分集めると、概ね30GB程度の分量になる。本稿を執筆した時点では、XBRLの全面導入から数年分しか経っていないが、20年分のデータが蓄積したならば500~600GB程度の分量となるだろう。ビッグデータというには小さいものであるが、これまでのパソコンではとうてい扱うことのできない分量である。

それゆえ、来るビジネス・アナリティックスの世界が到来した際に、我々研究者もビッグデータと格闘し、ディープラーニングのような手法を駆使して分析ができるように備える必要がある。本稿がそれに資することができれば幸いである。

(注)

- 1) これと同様のことは、近年のテキストマイニングを応用した研究にも散見される。どのようなアルゴリズムでテキストデータをスコアリングしているかについてまったく言及せず、統計ソフトに流し込んだ結果をただ単に示すだけでは、研究のクレディビリティは確保されているとは言えない。
- 2) 論理演算についての詳細は、ブール代数と論理演算に関する入門書、たとえば小倉・高浜 (1991) 等を参照されたい。
- 3) 多項ロジスティック回帰の場合は、活性化関数としてはシグモイド関数と関連が深いソフトマックス関数 (softmax function) が用いられる (橋本 2017, pp. 208-209)。また誤差関数 E は、対数尤度 $L(w)$ を用いて定義されるが (橋本2017, pp. 206-209)、これらの内容は紙幅の関係でここまでにとどめることにしたい。
- 4) 畳み込み演算およびプーリング演算の具体的な例は、たとえば斎藤 (2017, p. 208およびp. 219) の説明を参照されたい。
- 5) デロイトトーマツでは「Audit Analytics」、KPMGでは「Data & Analytics」、EYやPWCでは「デ

ータアナリティックス」といった表現で語られているが、いずれもビッグデータやAIを用いた分析により内部監査サービスの充実を謳っている。

(参考文献)

- Altman, E. I. (1968) "Financial Ratios, Discriminant Analysis and the Prediction of Corporate Bankruptcy," *Journal of Finance*, 23 (4) : 189-209.
- Fanning, K., Cogger, K., & Srivastava, R. (1995) "Detection of management fraud : a neural network approach," *International Journal of Intelligent Systems in Accounting, Finance & Management*, 4 (2) : 113-26.
- Fanning, K. & Cogger, K. (1998) "Neural network detection of management fraud using published financial data," *International Journal of Intelligent Systems in Accounting, Finance & Management*, 7 (1) : 21-41.
- Hoogs, B., Kiehl, T., Lacombe, C., & Senturk, D. (2007) "A genetic algorithm approach to detecting temporal patterns indicative of financial statement fraud," *Intelligent Systems in Accounting, Finance and Management*, 15 (1-2) : 41-56.
- Kotsiantis, S., Koumanakos, E., Tzelepis, D. & Tampakas, V. (2006) "Forecasting fraudulent financial statements using data mining," *International Journal of Computational Intelligence*, 3 (2) : 104-110.
- Loebbecke, J., Eining, M. and Willingham, J., (1989) "Auditor's experience with material irregularities : frequency, nature, and detectability," *Auditing : A Journal of Practice & Theory*, 9 : 1-28.
- McCulloch, Warren S. & Pitts, Walter (1943) "A logical calculus of the ideas immanent in nervous activity". *The bulletin of mathematical biophysics*, 5 (4) : 115-133.
- Minsky, M. & Papert, S. (1969) *Perceptrons : An Introduction to Computational Geometry*, The MIT Press.
- Ravisankar P., Ravi V., Rao G. R. & Bose I. (2011) "Detection of financial statement fraud and feature

- selection using data mining techniques,” Decision Support System, 50 (2) : 491-500.
- Rosenblatt, F. (1958). “The perceptron : a probabilistic model for information storage and organization in the brain”. Psychological Review, 65 (6) : 386-408.
- Rumelhart, D. E., Hinton, G. E. & Williams, R. J. (1986) “Learning representations by back-propagating errors,” Nature 323 (6088) : 533-536.
- Rumelhart, D. E. & McClelland, J. L. (eds.) (1986) Parallel Distributed Processing, The MIT Press.
- 小倉久和・高浜徹行（1991）『情報の論理数学入門—ブール代数から述語論理まで』近代科学社.
- 斎藤康毅（2016）『ゼロから作るDeep Learning—Pythonで学ぶディープラーニングの理論と実装』, オライリージャパン.
- 坂上学（2012）「会計情報の伝達媒体の変革」『企業会計』, 64 (1) : 53-58.
- 坂上学（2015）「データマイニングと会計研究」『企業会計』, 67 (2) : 104-107.
- 坂上学（2017）「会計研究におけるディープラーニングの適用可能性」『産業経理』, 77 (3) : 84-97.
- 坂上学・清水泰洋・島本克彦（2007）「『質問票調査』の分析結果—テキストマイニング分析」, 中野常男編著『複式簿記の構造と機能—過去・現在・未来』（第3章所収）, 同文館, 81-101.
- 白田佳子・坂上学（2008）「人工知能アプローチによる「継続企業の前提」の解析—テキストマイニングによる非会計情報の分析—」, 高田敏文編著『事業継続能力監査と倒産予測モデル』（第6章所収）, 同文館, 181-215.
- 西尾泰和（2015）『ルールベースから機械学習への道』, サイボウズラボ.
(<https://www.slideshare.net/nishio/ss-53221829>)
- 橋本泰一（2017）『データ分析のための機械学習入門—Pythonで動かし、理解できる、人工知能技術』, SBクリエイティブ.
- 富士通研究所（2012）「業界初！ビッグデータを簡単に利活用するための分析シナリオを自動的に推薦する技術を開発」
(<http://pr.fujitsu.com/jp/news/2012/08/21.html>)
- 法政大学環境センター（2016）『法政大学環境報告書』
法政大学.
- (<https://www.hosei.ac.jp/documents/ecoreport2016.pdf>)