

取締役推薦理由文を用いた取締役のスキル・マトリックス分類モデルの開発

山脇 大^{1,2} 野中 賢也³ 田村 光太郎³

高野 海斗¹ 中川 慧^{1,4}

¹ 野村アセットマネジメント株式会社 ² 北海道大学

³ 株式会社ユーザベース ⁴ 大阪公立大学

yamawaki.d.ai@gmail.com {kenya.nonaka,koutarou.tamura}@uzabase.com

{takaito0423,kei.nak.0315}@gmail.com

概要

株式会社のガバナンスの中核を担う取締役会の構成は、時代の変化に伴い透明性や多様性の確保が求められてきた。近年では、取締役会の機能強化を支える手段としてスキル・マトリックスが注目されている。しかし、その作成は各企業に委ねられ、客観性や標準化の欠如が課題となっている。そこで本研究では、取締役のスキル・マトリックス分類モデルの開発を行う。具体的には企業の公開している取締役推薦理由文からスキルを抽出・データセットを作成し、スキルを定義するとともに、スキル分類を多ラベル問題として定義する。その上で、ラベル不均衡に対応する学習フレームワークを提案し、その有効性を実証した。

1 はじめに

取締役が持つスキルや専門性を可視化し、取締役会の機能をより具体的に評価する手段として「スキル・マトリックス」が注目を集めている。米国において取締役のスキルは、2002年のSOX法以降に関心が高まり、上場企業に財務専門知識の開示を義務付けている[1]。日本においても、2021年6月のコーポレートガバナンス改定¹⁾により企業の取締役会のスキル・マトリックスの開示が推奨される等、制度的発展も進んでいる。企業はスキル・マトリックスを開示することによって、自社の取締役会が、経営方針や事業戦略等に照らして適切な知識・能力・経験等を備えており、十分な実効性を有することを伝えることができる。また、投資家等のステークホルダーからすると、開示されたスキル・マト

リックスより、企業の事業活動や経営戦略上、必要なスキルが適切に取締役会に備えられているのかを確認・判断することができる。そのため、取締役のスキル・マトリックスはベストプラクティスとして認識されている[2]。

その一方で、スキル・マトリックスの作成は各企業に委ねられていることもあり、依然として「見たい」を揃える範疇から出ることができず、実態を表していないケースも散見される。また、新任取締役の選任にあたっては、取締役会の現在のスキルセットと不足しているスキルを識別し、弱点を補うことが求められるが[3]、現状のスキル・マトリックスは定性的なデータが中心であり、評価の客観性や標準化が課題となっている。

このような背景を踏まえ、本研究では、企業が公開している取締役の選任に関するテキストデータを活用し、取締役のスキル・マトリックスを自動作成するモデルを開発する。具体的には、まずスキルとして認識する取締役のスキルセットの定義を行う。次に、株主総会招集通知や株主総会参考書類に記載された取締役推薦理由文と、そこから読み取れるスキルのラベルを付与したデータセットを作成する。そして、推薦理由文を入力として、該当する取締役のスキル(マトリックス)を出力する多ラベル分類モデルを構築する。なお、取締役が保有するスキルには偏りが存在することから、スキルごとの二値分類モデルを学習させる方法と、他ラベル分類モデルを学習させる方法の比較も行う。交差検証により、スキル・マトリックスの予測精度を実証的に検証する。

本研究の手法により、実際の推薦理由文からスキルを抽出し、取締役の経験等の実態に基づいたスキ

1) <https://www.jpx.co.jp/news/1020/20210611-01.html>

ル・マトリックスを作成することが可能となる。また異なる企業間でのスキル表現を統一的な形式に変換し、比較可能なデータセットを生成できる。さらに企業が明示していないスキルについても、推薦理由文の記述から関連するスキルを推定でき、現在の取締役会が備えているスキルと、理想的なスキルセットとの間のギャップを定量的に分析することが可能となる。

2 データセット

本研究では、2024年に開催された株主総会に先立って公表された各種の書類から作成したデータセットを用いる。我が国では会社法により取締役等の機関の選任および解任に関する事項は株主総会で(普通)決議が必要である。特に取締役の選任に関しては株主招集通知または株主総会参考書類において取締役推薦理由が記載されることが多い。これらに記載された取締役推薦理由文(以下、「推薦理由文」)を抽出してデータセットを作成する。

対象企業は流動性を考慮し、TOPIX500構成銘柄のうち、業種の偏りを踏まえて51社の企業を選定し、それら企業の株主招集通知または株主総会参考資料から推薦理由文を抽出した²⁾。そして、推薦理由文から読み取れる取締役が保有していると考えられるスキルを、経験豊富なアナリストの手によって付与した。結果、全500人の取締役の推薦理由およびスキルのデータセットを作成した。本研究で使用するスキルおよびその定義を付録の表2に示す。

3 問題設定

問題1(スキル・マトリックスの多ラベル分類問題)。推薦理由文を表すテキストデータ $x \in \mathcal{X}$ に対してスキル・マトリックスのラベル $y \in \{0, 1\}^K$ を予測するモデル $h_\theta : \mathcal{X} \rightarrow [0, 1]^K$ を学習する。ここで、 \mathcal{X} は推薦理由文が属するテキスト空間、 K はスキルの数、 θ はモデルのパラメータである。モデルのパラメータ θ を、トレーニングデータ $\mathcal{D} = \{(x_i, y_i)\}_{i=1}^N$ に基づいて損失関数 $\mathcal{L}(h_\theta, x, y)$ を最小化するように学習する。

$$\theta^* = \arg \min_{\theta} \mathbb{E}_{(x, y) \sim \mathcal{D}} \mathcal{L}(h_\theta(x), y). \quad (1)$$

多ラベル分類問題では、出力層にはシグモイド関

2) TOPIX500は、TOPIXの構成銘柄(プライム市場上場銘柄)から流動性や時価総額が一定基準を満たした500銘柄で構成される株式指数であり、TOPIXと比較して、より大型で流動性の高い銘柄が中心となる。

数を用いラベル k に対する確率は次のように定義される。

$$P(y_{ik} = 1 | x_i) = \sigma(h_k(x_i)) = \frac{1}{1 + \exp(-h_k(x_i))}, \quad (2)$$

ここで $h_k(x_i)$ はラベル k に対応するスコアを表す。また、損失関数として Binary Cross Entropy(BCE)を用いる。推薦理由文 x_i に対し、正解ラベルが $y_i = (y_{i1}, \dots, y_{iK})$ の場合、BCEは次のように定義される[?, ?]。

$$\mathcal{L}_{\text{BCE}} = -\frac{1}{NK} \sum_{i=1}^N \sum_{k=1}^K \left[y_{ik} \log P(y_{ik} = 1 | x_i) + (1 - y_{ik}) \log(1 - P(y_{ik} = 1 | x_i)) \right]. \quad (3)$$

4 提案フレームワーク

多ラベル分類問題における不均衡ラベルの影響を軽減するため、本研究では次の枠組みを採用する。

4.1 ラベルごとの独立モデル学習

前章の損失関数の下での不均衡ラベルの多ラベル問題の学習において、以下のスコア競合が生じる。すなわち、多数派ラベルに対応する損失が相対的に大きくなり、モデルのパラメータ更新が多数派ラベルに偏る。また、少数派ラベルのスコアの更新が、他の多数派ラベルのスコアの更新による影響を受ける[?]。この問題を緩和するため、本研究では以下のようにラベルごとの独立した二値分類器を構築し、それぞれ独立に最適化する。

$$h_k : \mathcal{X} \rightarrow \{0, 1\}, \quad k = 1, 2, \dots, K. \quad (4)$$

各ラベル k の損失関数は独立に定義される。

$$\mathcal{L}_k = -\frac{1}{N} \sum_{i=1}^N \left[y_{ik} \log P(y_{ik} = 1 | x_i) + (1 - y_{ik}) \log P(y_{ik} = 0 | x_i) \right]. \quad (5)$$

これにより、少数派ラベルのスコア更新が多数派ラベルの影響を受けることがなくなり、スコア競合を回避できる。

命題1. ラベルごとの独立した二値分類器を構築することで、スコア競合を回避することができる。

4.2 重み付き損失関数の導入

各ラベルで定義される損失関数において、ラベル正負の不均衡によるバイアスを緩和するため、各ラベル k の不均衡度に応じた重み w_k を損失関数に導

入する。具体的には、以下のような重み付き損失関数を使用する。

$$\mathcal{L}_{\text{W-BCE};k} = -\frac{1}{N} \sum_{i=1}^N \left[w_k y_{ik} \log P(y_{ik} = 1|x_i) + (1 - y_{ik}) \log P(y_{ik} = 0|x_i) \right] \quad (6)$$

ここで、 w_k はラベル k における正負の不均衡を是正するための重みであり、以下のように設定する。

$$w_k = \frac{P(y_k = 0)}{P(y_k = 1)} \quad (7)$$

本研究のデータセットのようなラベル間で不均衡度合いが大きく異なる多ラベル問題に対して、重み付きの二値分類をそれぞれ行うことで、式 (3) を損失関数に使用する多ラベル問題よりも精度の向上が見込まれる。

5 実証分析

5.1 実験設定

前章に記載した通り、与えられた推薦文に対し、12 種類のスキルラベルの付与を行う。提案手法となる (I) ラベルごとの独立モデル学習のほか、比較のための実験設定として、(II) 多ラベルとしての学習、より高度な文脈を理解できる (III) 大規模言語モデルの利用を検討し、下記のとおり学習・評価を行う。

I. ラベルごとの独立モデル学習

まず、ラベル分類の基盤モデルとして、BERT[4] を利用する。BERT は、transformer をベースとして、さまざまな基礎的な自然言語処理タスクで性能を発揮するエンコーダ型の言語モデルである。日本語テキストデータを扱うために、tokenizer を MeCab に変更し、日本語の wikipedia で事前学習を行ったものが東北大から公開 [5] されている。本研究では、この東北大 BERT の Large サイズを使い、モデルから出力される [CLS] トークンの埋め込み表現を扱う。

本研究で提案する 12 種類のラベルの付与は、3 章問題設定で述べたように、ラベルごとに二値分類行うモデル（つまり、各ラベルを付与するモデルが独立に 12 個できる）を構成することで行う。[CLS] トークンの埋め込み表現に線形層とシグモイド関数を加えることで $\{0, 1\}$ に変換する分類モデル h_k を構成している。推薦理由文とスキルのラベルがセットになっている作成したデータセットを用いて、ラベルごとにモデルを学習させる。具体的には、ある

ラベル k の付与を学習するためのデータセット D_k は、入力となる推薦理由文 x_i とラベル ky_{ik} の組となる。

II. 多ラベル問題としての学習

上記の BERT モデルにおける扱いにおいて、[CLS] トークンの埋め込み表現に線形層を加えたものから、12 種類のラベルの有無 $\{0, 1\}^K$ 、ここで $K = 12$ に変換する分類モデルを構成して、多ラベル問題として付与するモデルを構成する。

なお、損失関数については、式 6 をラベルについて足しあげたものを利用する。

$$\mathcal{L}_{\text{W-BCE}} = \frac{1}{K} \sum_k \mathcal{L}_{\text{W-BCE};k} \quad (8)$$

III. 追加的実験

文脈をより高度に理解できる大規模言語モデルを使い、スキルラベルを分類するタスクへの適用を検討する。ここでは、(A)gemma-2-2b-jpn-it[6] と (B)GPT-4o[7] に対して、下記のように推薦理由文を入力として、そのスキルラベルの付与を行った。

インストラクション

以下の経歴・背景をもつ人物が [ラベル名] に専門性があると判断できれば 1 を、そうでなければ 0 を出力してください。

[推薦理由文]

A. gemma-2-2b-it

gemma-2-2b-it では、LoRA を利用して、推薦理由文を上記インストラクションとともに学習させる。したがって、ラベルごとの独立モデル学習となるため、こちらでも独立したモデルが 12 個できる。

B. GPT-4o

OpenAI の GPT-4o に関しては、上記のプロンプトを与えて、ラベルごとに 0, 1 を答えさせる Zero-shot 分類を行う。

5.2 評価方法

スキルラベルは表 2 にあるように、正負比率に偏りがあるため、ラベルの正負比率が保存するようにデータを 5 分割し、層化抽出を行って学習と評価を行う。つまり、ここで 5 分割したデータのうち、4 つを訓練に、1 つをテストにして、5 パターンのテストセットにおける精度を算出することで、モデルの精度や問題設定などの考察を行っている。

多ラベルとして扱う方式 (II) においても層化抽出

表 1: 数値実験結果

スキル	(I)	(II)	(III).A	(III).B
企業経営	0.96	0.89	0.89	0.91
法務・コンプライアンス	0.73	0.72	0.56	0.67
リスク管理	0.71	0.70	0.47	0.37
財務・会計・税務	0.79	0.76	0.76	0.71
海外経験	0.87	0.85	0.81	0.74
人事・労務	0.86	0.85	0.90	0.45
技術・専門性	0.80	0.84	0.65	0.68
研究開発	0.67	0.54	0.67	0.48
営業・マーケティング・企画・開発	0.65	0.62	0.62	0.63
行政・政府機関	0.79	0.80	0.75	0.79
金融	0.85	0.87	0.96	0.72
ESG・サステナビリティ	0.75	0.76	0.67	0.66
Micro	0.84	0.82	0.78	0.78
Macro	0.79	0.77	0.73	0.66

[8] を使い、同様に 5 分割する交差検証を行う。

各実験設定での精度指標は、交差検証における各テストセットを統合して算出した F1 値とする。F1 値は、混同行列 (confusion matrix) の、各要素である真陽性 (TP: True Positive), 真陰性 (TN: True Negative), 偽陽性 (FP: False Positive), 偽陰性 (FN: False Negative) の 4 つの頻度情報を用いて以下のように計算することができる。

$$\text{適合率} = \frac{TP}{TP+FP} \quad (9)$$

$$\text{再現率} = \frac{TP}{TP+FN} \quad (10)$$

$$\text{F1 値} = \frac{2 \times \text{適合率} \times \text{再現率}}{\text{適合率} + \text{再現率}} \quad (11)$$

5.3 実験結果と考察

評価結果を表 1 に示す。12 種類の各ラベルごとの付与判定モデルを構築する方式 (I) により、全体の精度が向上することが分かった。特に、スキル評価において重要と考えられる「企業経営」「財務・会計・税務」「ESG・サステナビリティ」の 3 つのラベルに注目すると、「企業経営」「財務・会計・税務」においては提案手法が優位となり、前者においては高い精度でラベルの付与がなされた。「ESG・サステナビリティ」においては、方式 (II) が精度で勝るが、同水準となった。

定性的には、「企業経営」「財務・会計・税務」において、複数のスキルに関連する単語や表現が多数出現するようなケースや、「金融」「財務・会計・税務」のように意味が近いラベル同士を的確に振り分けるケースにおいて、方式 (I) が優位となるケースが多かった。

「ESG・サステナビリティ」では、全体として、製

品や事業計画名に表れる表現（脱炭素やクリーンなど）に反応し、ラベルが付与されているとみられ、より文脈に則した付与が必要と思われる。

一方で、より文脈を考慮する大規模言語モデルを使う方式 (III) では、全体的に精度のばらつきが大きいことと、「法務・コンプライアンス」「リスク管理」や「技術・専門性」「研究開発」といった意味が近いスキルラベルの間での付与の誤りが多く、学習やプロンプトエンジニアリングの再考が必要である。

これらのことから、入力テキストにおける異なるラベルを同時的に扱うことや、ラベル間の差異が小さいケースにおいて提案手法である方式 (I) が優位に現れる傾向にあると考えられる。本研究のように、スキルのラベルが 12 と少ない、学習データの作成にコストがかかるため学習データが少量である、ラベルごとに出現が不均衡である、といった条件の下で方式 (I) は有効である。一方でラベルごとの独立モデル学習は、ラベルの数と学習時間や推論時間が線形関係となるため、扱いたいラベル数が多くなると運用が難しい。また、一般的には大量に質の高い学習データがある場合、マルチタスクの学習を行ったモデルが、独立モデルよりも精度が良くなるケースも多い。したがって、初期段階では精度重視で独立モデルを学習させつつ、データ拡張を行うことで、方式 (II) のモデルを学習させるといった、段階的なモデル開発が重要である。

6 まとめと今後の展望

本論文の主な貢献は、株主総会招集通知から取締役のスキルをラベル付けした新データセットの構築、推薦理由文を基にしたスキル・マトリックスの多クラス分類問題の定義、そして不均衡データセットにおける分類精度向上を目的とした OVA フレームワークの提案と実証である。

今後の課題は、業種ごとの役員スキル・マトリックスのカバレッジや偏りを分析し、どのようなスキル構成の役員メンバーが形成されているかを、またそれらの特徴がガバナンスや企業収益性等に与える傾向を定量的に検証することである。

参考文献

- [1] Udi Hoitash, Rani Hoitash, and Jean C Bedard. Corporate governance and internal control over financial reporting: A comparison of regulatory regimes. *The accounting review*, Vol. 84, No. 3, pp. 839–867, 2009.
- [2] Nemmara K Chidambaran, Yun Liu, and Nagpurnanand Prabhala. Director diversity and inclusion: At the table but in the game? *Financial Management*, Vol. 51, No. 1, pp. 193–225, 2022.
- [3] Catherine M Dalton and Dan R Dalton. Boards of directors: Utilizing empirical evidence in developing practical prescriptions. *British Journal of management*, Vol. 16, pp. S91–S97, 2005.
- [4] Chang Ming-Wei Lee Kenton Devlin, Jacob and Kristina Toutanova. BERT: Pre-training of deep bidirectional transformers for language understanding. In *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers)*, pp. 4171–4186, June 2019.
- [5] (2024-12 閲覧) . <https://huggingface.co/tohoku-nlp/bert-large-japanese-v2>.
- [6] (2024-12 閲覧) . <https://huggingface.co/google/gemma-2-2b-jpn-it>.
- [7] (2024-12 閲覧) . <https://chatgpt.com/>.
- [8] Konstantinos Sechidis, Grigorios Tsoumakas, and Ioannis P. Vlahavas. On the stratification of multi-label data. In *Machine Learning and Knowledge Discovery in Databases - European Conference*, Vol. 6913, pp. 145–158, 2011.

A 取締役のスキル定義および保有判定の例示

推薦理由文の例示として以下に、上場株式会社の株主招集通知から取得可能な推薦理由文の典型的な様式例を記載する。

推薦理由文の例示

取締役候補者とした理由および期待する役割:

同氏は、国内外の大手金融機関 XX や事業会社 XX で役員としての要職を歴任し、当社入社後は主要な経営ポジションにおいて指導的役割を担い、XXXX 年 XX 月より当社取締役社長をつとめております。このような豊富な経験と金融に関する高度な専門性にに基づいた、当社の経営全般の適切な監督と意思決定を期待すべく、同氏を引き続き取締役候補者として推薦いたします。

本研究では、推薦理由文から下記の内容に関しての記載があった場合に該当する個人が、そのスキルを保有していると判定する。スキルは個人に対して、複数保有していることも考慮する。

企業経営: 会長、社長、頭取、最高経営責任者 (CEO)、取締役、執行役員などの役職で培われた経営全般の経験。

法務・コンプライアンス: 法務部門での経験、弁護士資格の保有。

リスク管理: リスク管理部門での経験。

財務・会計・税務: CFO、監査法人での実務経験、公認会計士資格や証券アナリスト資格の保有。

海外経験: 海外駐在勤務、多国籍企業や国内企業の海外事業部門における経験。

人事・労務: 人事部門での管理職経験。

技術・専門性: エンジニアリング部門や大学・研究機関での経験、事業関連分野への精通。

研究開発: R&D や新規事業部門での経験。

営業・マーケティング・企画・開発: 営業・マーケティング部門や企画・開発部門での経験。

行政・政府機関: 官公庁・公的機関での経験。

金融: 銀行、証券会社、運用会社等での経験。

ESG・サステナビリティ: ESG・サステナビリティに関する知識・経験。

表 2 が本研究で使用するスキルラベルデータにおける付与件数である。

表 2: 本研究で使用するスキル (ラベル) および定義とデータにおける付与件数

スキル	件数
企業経営	340
法務・コンプライアンス	35
リスク管理	35
財務・会計・税務	83
海外経験	115
人事・労務	34
技術・専門性	201
研究開発	18
営業・マーケティング・企画	81
行政・政府機関	34
金融	51
ESG・サステナビリティ	91
	455

B 命題の証明

証明 多ラベル分類問題において、各ラベル k に対して独立した二値分類器 $h_k(x)$ を学習する場合と、1 つのモデルで一括で学習する場合を比較する。一括して学習する場合、モデルパラメータ θ を共有するため、(3) 式の損失勾配は全ラベル k に対して合算され、次のように書ける。

$$\frac{\partial \mathcal{L}_{\text{BCE}}}{\partial \theta} = \sum_{k=1}^K \frac{\partial \mathcal{L}_k}{\partial \theta}. \quad (12)$$

これにより、多数派ラベルの影響が強く反映され、スコア競合が発生する。一方で、各ラベルごとに学習する場合、各ラベル k に対して次の損失関数を個別に最適化する。

$$\mathcal{L}_k = -\frac{1}{N} \sum_{i=1}^N \left[y_{ik} \log P(y_{ik} = 1 | x_i) + (1 - y_{ik}) \log(1 - P(y_{ik} = 1 | x_i)) \right]. \quad (13)$$

各ラベル k に対する確率 $P(y_{ik} = 1 | x_i)$ は他のラベル $j (j \neq k)$ に依存しないため、勾配更新が完全に独立して行われる。そのため、多数派ラベル k のスコア $h_k(x)$ が他のラベルの損失勾配に影響を与えずスコア競合は発生しない。 □