

2018年12月12日 全9頁

# テキストマイニングによる統合報告書評価

## 重要語抽出により定性情報の客観評価が可能に

調査本部 小林 若葉  
リサーチ業務部 大和 敦

### [要約]

- 本稿では、重要語の抽出を行うツールである TF-IDF 分析というテキストマイニングの手法を用いて企業が発行する統合報告書の客観的評価を行った。
- 一つ目の分析では同業種内の他社比較において個社特有の重要語が、二つ目では同一企業の経年比較において当該年特有の重要語が抽出されることが確認できた。得られた重要語を手掛かりに統合報告書を読むことで、企業が注力していることを読み取ることが可能である。
- TF-IDF はテキストデータの特徴をつかむには有用な手段だといえる。技術的な課題は残るものの、重要語抽出によるテキストマイニング分析の活用方法を見出せたことは大いに価値がある。

## 統合報告書の開示拡大と客観分析の必要性

企業による非財務情報開示の動きが加速している。世界的に気候変動や資源の枯渇といった環境問題、労働者の人権問題などの社会問題への関心が高まる中で、2006年に国連責任投資原則(PRI)が公表され、ESG(環境・社会・ガバナンス)投資が広がるきっかけとなった。ESG投資は環境問題、社会問題、そして企業統治の問題が運用パフォーマンスに影響を及ぼしかねないという理由から、これらの課題に配慮した企業経営を行う企業に投資することで、長期的リターンを追及するという考え方である。2007年以降に欧米から発生した一連の金融危機も、金融市場における短期的な利益追求の志向が金融危機を招いたという認識から、企業の長期的な成長と企業価値拡大を評価するESG投資を後押しした。

日本においても、世界最大規模の機関投資家である年金積立金管理運用独立行政法人(GPIF)が2015年にPRIに署名し、2017年には三つのESG指数(総合型としてFTSE Blossom Japan IndexとMSCI ジャパン ESG セレクト・リーダーズ指数、テーマ型(社会(S))としてMSCI 日本株女性活躍指数)を選定し、これらに連動したパッシブ運用を開始したことで国内におけるESG投資を加速させた。

2015年には国連総会で「持続可能な開発目標（SDGs）」が採択された。SDGsは同年に期限を迎えた「ミレニアム開発目標（MDGs）」を前身とするが、開発途上国向けの目標であったMDGsと異なり、SDGsを構成する17の目標には技術革新、労働者の働きがい、気候変動など、先進国の課題も含まれており、全世界で対処すべき課題とされている。SDGsにおける企業の役割も大きく、企業がESGに配慮した取り組みを行うにあたり、目指すべきゴールが明示されたといえる。他方、SDGsは投資家にとってもESG投資を行うための企業評価の共通言語となると期待される。

このようなESG投資の動きの中で企業の非財務情報が重要視されるようになり、財務情報と非財務情報の両方を盛り込んだ「統合報告書」が注目されるようになった。日本において統合報告書の発行企業数は年々増加しており、2017年には341社（KPMGジャパンの調査）と、2016年から50社増加している。しかし、有価証券報告書等とは異なり、統合報告書には定型様式や開示項目を規定するものがないため、その内容や分量は様々であり、各企業の創意工夫が見られる。加えて、企業の利益や株主資本、配当金といった財務情報は数値やグラフで示すことができる一方で、ESGに配慮した取り組みの紹介や中長期経営計画といった非財務情報は定性的な情報となりやすい。実際に数十ページから数百ページからなる報告書を複数社分読み、比較することは困難であることから、今後さらに重要となる非財務情報を効率的に、そして客観的に読む必要性が増してくると考えられる。

## テキストマイニングを用いた重要語抽出による分析

GPIFは2018年1月に「GPIFの国内株式運用機関が選ぶ『優れた統合報告書』と『改善度の高い統合報告書』」を発表した。GPIFは、国内株式運用機関16社に対し、優良な統合報告書発行企業を最大10社ずつ選定するよう依頼をしており、「優れた統合報告書」で延べ70社、「改善度の高い統合報告書」で延べ68社が選ばれた。そのうち、「特に多くの運用機関から高い評価を得た『優れた統合報告書』」（以下、特に優れた報告書）として味の素、コニカミノルタ、オムロン、伊藤忠商事、丸井グループの5社、「特に多くの運用機関から高い評価を得た『改善度の高い統合報告書』」（以下、特に改善度の高い統合報告書）として大和ハウス工業、住友金属鉱山、オムロン、住友商事の4社を取り上げ、各企業の統合報告書が評価された理由についてのコメントを紹介した。

本稿では、人の目による主観的な判断に基づいて評価された統合報告書が、機械的な観測手法を用いたときに、果たして同様の評価になるのかを確かめたい。用いる手法は、プログラミング言語のPythonによる、大量のテキストデータから情報を取り出すテキストマイニングであり、具体的には、テキストにおける重要語の抽出を行うツールのTF-IDFである。TF-IDFは情報検索などでよく用いられる手法であり、定性情報を多く含む統合報告書の内容の特徴をつかむのに適している。テキスト内で「出現回数の多い単語」を拾い、「出現回数の多い単語」ほど「重要な単語」であると評価するものである。しかし、出現回数が多い単語の中には「こと」「とき」など、多くの文書で出現する「一般語」も含まれてしまうため、これは取り除く。すなわち、

TF-IDF 分析では、

$$tf = \frac{\text{文書 A における単語 X の出現回数}}{\text{文書 A におけるすべての単語の出現回数の和}}$$

$$idf = \log \frac{\text{総文書数}}{\text{単語 X が出現する文書数}}$$

として、TF-IDF の値を次のように定義する。

$$TF-IDF = tf \times idf$$

つまり、単語 X の出現頻度  $tf$  に「逆文書頻度」と呼ばれる、多くの文書に出現すればするほど値が小さくなる  $idf$  を掛けることで求められる。これにより、文書の中で出現回数が多く、かつ他の文書の中であまり出現しない単語が抽出される。

ここでは TF-IDF を用いて二つの分析を行う。第一の分析は、同一業界内の企業の中で GPIF の運用機関に選ばれた「優れた統合報告書」と選ばれなかった統合報告書についての比較分析、第二は同一企業における 2017 年版の「改善度の高い統合報告書」とそれ以前の統合報告書についての比較分析である。TF-IDF により出現回数が多く、かつ他の統合報告書の中であまり出現しない単語を抽出し、これを重要語として、複数の統合報告書の内容を定量的に、また効率的に把握しようとするものである。

## 同業他社比較

第一の分析に用いた統合報告書発行企業は、日本の上場企業で主な投資対象となると考えられる TOPIX500 の構成銘柄に含まれる 2017 年に統合報告書を発行した食品企業とした。この条件に適合する企業は味の素、キリンホールディングス、伊藤園、アサヒグループホールディングス、カゴメ、サッポロホールディングスの 6 社であった。さらにこの 6 社を三つのグループに分ける（図表 1）。一つ目は「特に優れた統合報告書」の味の素、二つ目は「優れた統合報告書」のキリンホールディングスと伊藤園、三つ目は「優れた統合報告書」には選ばれなかったアサヒグループホールディングス、カゴメ、サッポロホールディングスである。それぞれをグループ A、グループ B、グループ C として、各グループの統合報告書を比較する。

図表 1 分析対象の統合報告書発行企業グループング

グループ	グループ名	企業名	ページ数	文字数
グループA	特に優れた統合報告書	味の素	84	64,811
グループB	優れた統合報告書	キリンホールディングス	90	74,447
		伊藤園	56	65,897
グループC	その他	アサヒグループホールディングス	82	61,576
		カゴメ	98	76,296
		サッポロホールディングス	70	72,424

（出所）年金積立金管理運用独立行政法人「GPIF の国内株式運用機関が選ぶ『優れた統合報告書』と『改善度高い統合報告書』」（平成 30 年 1 月 19 日）等より大和総研作成

下準備として各社のウェブサイトから統合報告書の PDF をダウンロードし、Python ライブラリの PDF Miner を用いてテキスト化を行う。得られたテキストデータを用いて TF-IDF により 6 社の統合報告書における各グループの統合報告書の重要語を抽出し、社名や冊子タイトル、意味をなさない単語を「不要語」として取り除いた結果（上位 25 位）が図表 2 である。

## 重要語の解釈

このグループ A の重要語の最上位には“ASV”という単語がある。“ASV”の意味を味の素の統合報告書で確認してみると、“Ajinomoto Group Shared Value”の略称であり、「社会課題の解決を通じて創出された経済価値を事業活動（技術、人財等）へ再投資することで、更なる社会課題の解決に貢献するという好循環（ASV サイクル）を作り出し、サステナブルな成長を実現するための戦略的な取り組み」だとしている。つまり、事業を通じて社会課題を解決し、収益を得ることで、社会価値と経済価値の創出を同時に行うという考えである。同報告書では、社長メッセージや社長と大学教授との対談で、同社における ASV の重要性を説明するとともに、従業員ひとりひとりが ASV の考え方を理解し、実践する組織風土づくりのために新設した、ASV を体現した事業を表彰する「ASV アワード」において大賞となった取り組みを紹介している点が特徴的である。

図表 2 「不要語」除去後の各グループの統合報告書の重要語と TF-IDF 値

グループA			グループB			グループC		
No.	重要語	TF-IDF値	No.	重要語	TF-IDF値	No.	重要語	TF-IDF値
1	ASV	438	1	ティー	330	1	取組み	486
2	2020年度	210	2	お茶	306	2	事業	428
3	アミノ酸	168	3	茶	300	3	アサヒスーパードライ	354
4	サイエンス	168	4	ライオン	276	4	トマト	334
5	ウインザー	156	5	のれん等償却	240	5	国際	249
6	スペシャリティ	147	6	価値創造基盤	222	6	欧州事業	246
7	ウインザー社	144	7	茶産地	222	7	ブランド	215
8	事業	137	8	クラフトビール	216	8	野菜	212
9	アミノサイエンス	132	9	お客様	206	9	酒	210
10	バイオ	120	10	事業	204	10	健康寿命	204
11	冷凍食品	114	11	生茶	198	11	お客さま	204
12	activity	108	12	協和発酵キリン	192	12	ビール	203
13	事業利益	108	13	バイオケミカル	186	13	売上収益	198
14	GROW	102	14	主義	186	14	カルビス	198
15	うま味調味料	96	15	一番搾り	180	15	農業	198
16	先端バイオ	96	16	茶産地育成事業	180	16	企業情報	198
17	味の素ウインザー社	96	17	CSV	178	17	成長	189
18	グローバル	90	18	ビール	176	18	菜園	186
19	向け	90	19	ブラジルキリン	174	19	商品	176
20	ファイン	90	20	茶殻	174	20	レモン	174
21	栄養	88	21	ティーテイスター	162	21	延伸	174
22	FIT	84	22	ミャンマー・フルフリー	150	22	国内	172
23	動物栄養	84	23	総合飲料事業	150	23	農事業	156
24	コーポレート	84	24	のれん	144	24	取組み	154
25	中計	81	25	ブランド	140	25	恵比寿	150

（出所）各社統合報告書より大和総研作成

実際に同報告書を「優れた統合報告書」に選んだ運用機関からのコメントでは、「トップメッセージの中で、主な経営指標が財務（経済価値）と非財務（社会価値）に関連付けて明確に説明されており、統合（ブランド価値）を生み出すというシナリオが明快」<sup>1</sup>と、ASV の概念の説明が評価されており、TF-IDF によって得られた“ASV”の重要性が高いという結果に合致している。

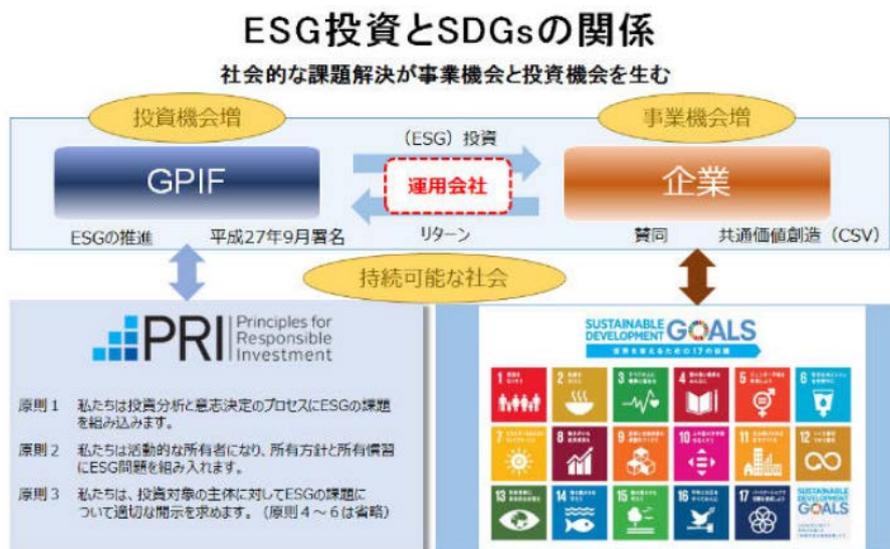
“ASV”は、企業の競争戦略論を専門とするアメリカの経営学者マイケル・E・ポーターハーバード大学教授が 2011 年に論文により提唱した“CSV”（Creating Shared Value、共通価値の創造）と味の素の創業以来の志が似ているとして、これに因んで名づけられた味の素独自の経営戦略である。CSV は経済的価値を創造しながら社会的ニーズに応えることで、社会的価値も創造するというアプローチであり、同氏は従来の事業との相関性が低い慈善活動的な CSR から、ビジネスとして社会課題を解決する CSV への転換を唱えた<sup>2</sup>。企業による社会的活動は一般に儲からないと捉えられてきたが、サステナブルに社会価値を提供するためには事業活動に結び付けることが重要だという考えから、近年 CSV が注目されている。GPIF も、Web サイトにおいて、GPIF が行う「ESG 投資と SDGs のつながり」に関し、「SDGs に賛同する企業が 17 の項目のうち自社にふさわしいものを事業活動として取り込むことで、企業と社会の『共通価値の創造』（CSV）が生まれ」、その取り組みによって企業価値が持続的に向上することで、長期的な投資リターンの拡大につながるとして、SDGs における社会課題を企業が事業活動として取り組むことを期待している（図表 3）。

これを踏まえ、グループ B の結果を見てみると、重要語の 17 位に上述の“CSV”があることが確認できる。キリンホールディングスでは、事業に関連が深い「健康」「地域社会」「環境」を中長期的に取り組む重点的な社会課題として、顧客の食生活を支える商品ラインアップや情報の提供を通じて生活習慣病への対処に貢献する、日本産ホップの品質向上と安全調達に取り組むとともに生産地域の活性化に寄与する、といった 17 の CSV コミットメントを宣言している。他方、伊藤園は CSR を幅広い「社会対応力」ととらえ直し、本業を通じた CSR に切り替えているという。同社は強みを活かせる、「環境」「消費者」「コミュニティ」の 3 分野に資源を重点配分し、社会課題解決と経済価値の同時実現に向けて CSV を実践している。

<sup>1</sup> 年金積立金管理運用独立行政法人（2018）「GPIF の国内株式運用機関が選ぶ『優れた統合報告書』と『改善度高い統合報告書』」（平成 30 年 1 月 19 日）

<sup>2</sup> 一方で、本来 CSV は CSR の代替とはならないという議論も存在する。ニッセイ基礎研究所の川村雅彦氏は、「日本における CSR・CSV・ESG—そして統合報告の現状認識～グローバルにビジネスの枠組が変わり始めた～」(ディスクロージャー&IR 総合研究所「ディスクロージャーニュース」、2016. 4、vol. 32、pp. 151-161)において、『『本来の CSR』では、自社事業の中に社会的課題の原因が内在すると認識し、それが社会や環境に及ぼす影響に対して、“本業＝自社事業のプロセスとプロダクト”において解決する責任があると考え」、他方、「CSV は、自社事業と直接的な関係はないものの、社会全体が抱える問題・課題を解決するために、“本業＝自社の強みを活かしたプロダクトの開発・販売”に挑戦することである」としている。故に同氏はポーター教授が「受動的 CSR」と表現した、本業外の寄付やフィランソロピーといった活動は「本来の CSR」とはそもそも別物であり、「CSR の実践」と「CSV の実現」の同時性が必要であると主張している。

図表3 GPIFのESG投資とSDGsの関係



(出所) 年金積立金管理運用独立行政法人 Web サイト

一方、グループ C の TF-IDF 結果を見てみると、“CSV”に関連した語が上位 25 語に入っていない。3 社の統合報告書を確認してみると、各社ともに CSR や ESG の取り組みを紹介してはいるものの、CSR 活動と事業との関わりは明確に示されていない印象だ。中には慈善活動的ではない事業を通じた社会価値創造に結びつくような取り組みもある。しかし CSV の考え方やそれを自社で実践することの重要性を十分に説明できていないために、取り組みの意味や生み出される経済価値が連想されにくいのではないだろうか。経済価値を生み出しながら社会価値を創造する CSV の考え方は社会に定着しているとは言いがたく、CSR や ESG の取り組みが必ずしも経済価値を生むと考えられているとは限らない。その中で CSV の概念を説明し、自社の事業にどのように組み込んでいき、どのような価値を創造するのかを丁寧に説明することが求められていると考えられる。経済価値と社会価値の創造過程を明示するというところにこそ、財務情報と非財務情報を合わせた統合報告書を発行する意義があるといえる。

## 経年比較と重要語解釈

第二の分析では「特に改善度の高い統合報告書」である大和ハウス工業の 2017 年版の統合報告書をそれ以前のもの（2013～16 年版）と比較する。第一の分析と同様に統合報告書 PDF のテキスト化、重要語の抽出と不要語の除去を行った後に残った語の一覧が図表 4（2013 年版、14 年版は省略）である。

図表4「不要語」除去後の各年次の統合報告書の重要語と TF-IDF 値

2015年版			2016年版			2017年版		
No.	重要語	TF-IDF値	No.	重要語	TF-IDF値	No.	重要語	TF-IDF値
1	事業	292	1	事業	208	1	事業	322
2	住宅	157	2	住宅	91	2	戸数	175
3	人権	108	3	成長	76	3	住宅	153
4	開発	101	4	開発	67	4	ものづくり	125
5	総合	100	5	社会	65	5	データ	120
6	技術	75	6	商品	64	6	基盤	113
7	啓発	70	7	お客さま	64	7	岩崎	100
8	団地	68	8	株主	60	8	監査役	99
9	戸数	68	9	賃貸住宅	60	9	平田監査役	95
10	実績	67	10	中期経営計画	56	10	お客さま	93
11	環境	65	11	経営	54	11	当社	90
12	流通	65	12	ハート	53	12	経営	89
13	ランキング	60	13	ニーズ	49	13	取締役会	88
14	社会	58	14	5次	48	14	25点	85
15	サービス	56	15	価値創造	48	15	開発	84
16	不動産投資	55	16	人財	45	16	向上	83
17	工事	55	17	社会的	45	17	価値創造	80
18	施設	53	18	5次中期経営計画	43	18	株主	79
19	お客さま	52	19	価値	43	19	技術	79
20	排出	52	20	売上高	42	20	社会	77
21	建設	50	21	取締役	41	21	労働時間	75
22	企業	50	22	目標	41	22	人財	74
23	賃貸住宅	50	23	環境	40	23	岩崎監査役	70
24	地域	50	24	拡大	40	24	価値	69
25	取締役	48	25	向上	40	25	自己資本	65

(注) 2013年版、2014年版は表記を省略。

(出所) 大和ハウス工業統合報告書(2013~2017年版)より大和総研作成

2015、16年版と比較して2017年版の統合報告書で特徴的なのは「岩崎監査役」と「平田監査役」という語が多く出現している点である。実際に同報告書で確認してみると、常勤監査役の平田氏と社外監査役の岩崎氏の2人の監査役対談が目玉の一つとして設けられていることがわかる。取締役会において牽制機能を果たす監査役会の重要性を説くと同時に、同社の取締役会における自由闊達な議論を評価しており、監査役から見たガバナンスの透明性を示唆する内容となっている。

もう一つ、目に留まる語として、2017年版の報告書の14位の重要語である「25点」を取り上げたい。同報告書内では、「データ編・非財務ハイライト」のページを設け、「男性育児休業の取得率」が5点/25点、「売上高当たりCO2排出量の2005年度比削減率」が19点/25点といったように、同社で設定したCSR課題の20指標について、2016年度実績を自己評価している。2016年版の報告書でもCSR課題についていくつかの取り組み目標と実績の数値をそれぞれ記載してはいたが、2017年版ではこれらを「指標」として一覧にし、点数までつけたことは画期的であると評価できる。

同報告書を「改善度の高い統合報告書」に選んだ運用機関からのコメントを見てみると、「監査役対談では経営の意思が現場にどのように浸透しているかという観点で議論が行われ、また不動産開発投資に関するリスク管理の開示も投資判断の観点で重要な情報。CSR自己評価は興

味深い」<sup>3</sup>としている。二つ目の分析でも、運用機関の評価コメントの内容を TF-IDF 分析により機械的に捕捉することができたといえよう。コメントから、目新しい監査役の対談に加えて、CSR 目標をスローガン倒れにさせず、自社の取り組みを振り返り、目標達成に邁進しようという決意が評価されているとわかる。同社のように、CSR の取り組みについて、定性情報だけでなく KPI（重要業績評価指標）を設定する企業が増加している。客観的に目標の達成度が見えることで、取り組みに対する評価がされやすくなるとの考えからであろう。

## まとめ

本稿では TF-IDF 分析というテキストマイニングの手法を用いて企業が発行する統合報告書の内容の客観的評価を行った。二つの分析のうち、一つ目では同業種内の他社比較において個社特有の重要語が、二つ目では同一企業の経年比較において当該年特有の重要語が抽出されることが確認できた。そして得られた重要語を手掛かりに統合報告書を読むことで、企業が注力していること、特に伝えようとしていることを読み取ることができ、GPIF の運用機関が同様の点の評価していることも分かった。すべての企業について同様とは限らないが、優良な統合報告書の重要語の特徴から、事業と CSR 活動のつながり、すなわち社会価値と経済価値を同時に創出しているということを明示すること、また、CSR 活動について KPI を設定し、達成状況を公表することが好評価を得る一因になるといえる。

一方、TF-IDF 分析の課題としては、以下の三つが挙げられる。第一の課題は、テキストの分量が多い文書ほど、出現回数の多い単語が重要語として評価されやすくなってしまふ点である。TF-IDF では出現回数の多い単語を拾い、他の文書にも出現する単語の TF-IDF 値を引き下げるが、他の文書にも出現する単語であったとしても、テキスト量が多い文書ほど、出現回数が比較的多い単語であれば TF-IDF 値は大きくなり、重要語とみなされてしまう。統合報告書は各企業によって分量が異なるため、TF-IDF においてテキスト量が多い統合報告書の重要語は過大評価され、テキスト量の少ないものの重要語は過小評価される可能性がある。

第二に、CSR 関連の単語を多用していたり、他社との差別化を図った独自の単語を作成・記載している報告書が、取り組み実態に関わらず優良な統合報告書として判断され得る点である。CSR 関連の単語を散りばめれば統合報告書としての質が高くなるということはない。ESG に関して自社として何をすべきなのかを明示し、実際の取り組み状況や事例を紹介することが統合報告書として本来あるべき姿ではないだろうか。

第三に、重要語が一覧としてアウトプットされても、単語だけでは何を表すのか判断できないという点である。例えば上述の大和ハウス工業の 2017 年版統合報告書における「25 点」という語は、それ自体には重要な意味があるわけではなく、これだけを見ても報告書の内容を評価

<sup>3</sup> 年金積立金管理運用独立行政法人（2018）

することはできない。内容を把握するためには「25点」をキーワードとして一度報告書で確認する必要がある。

第二、第三の課題の今後の改善策としては、テキスト内の単語間の関連性を見るため、単語間の距離を測る方法がある。大和ハウス工業の統合報告書においては、「25点」という語の周辺には「CSR 自己評価指標」「目標」「実績」といった語が散見される。もしこれらの語の間の距離が近いと判断され、関連性が見える化されれば、「25点」という、これだけでは見落としてしまいがちな重要語の解釈がより容易になると考えられる。これにより CSR に関連する単語の量だけでなく、ある程度内容を把握したうえで報告書の評価もできるだろう。

とはいえ、大量のテキストデータから重要語を抽出することが、統合報告書の評価に役立つことは分かった。本稿で行った一つ目の分析では6社分、二つ目の分析では5年分の統合報告書を分析対象としたが、これらの報告書の一つ一つ読んで評価を行うのは非効率であり、主観的な判断にもなり得る。統合報告書における重要語が得られれば、これを道しるべとし、同報告書で確認することでこれを客観的に評価することが可能となる。TF-IDF はテキストデータの特徴をつかむには有用な手段だといえる。技術的な課題は残るものの、重要語抽出によるテキストマイニング分析の活用方法を見出せたことは大いに価値があろう。