

# ニューステキストを用いた ESG ファクター運用

## ESG factor investing using news text data

秋山祥伍<sup>1\*</sup> 江口潤一<sup>2</sup> 鈴木智也<sup>1,2</sup>  
Shogo Akiyama<sup>1</sup> Junichi Eguchi<sup>2</sup> Tomoya Suzuki<sup>1,2</sup>

<sup>1</sup> 茨城大学大学院理工学研究科

<sup>1</sup> Graduate School of Ibaraki University

<sup>2</sup> 大和証券資信託委託株式会社

<sup>2</sup> Daiwa Asset Management Co.Ltd.

**Abstract:** ESG investment is getting popular as an investing method that evaluates environmental, social and governance (ESG) efforts of each company. However, although the disclosure of information on ESG is already common, it is difficult to objectively evaluate ESG efforts of companies due to the ambiguity in the definition of ESG words: environment, society, and governance. For this reason, we applied Word2Vec to extract similar words to each ESG word, and evaluated ESG efforts of companies from the viewpoint of these extracted words. As a result, we confirmed that the companies with higher ESG score can make a more profitable portfolio than those with lower ESG score, and this ESG score can be useful for factor investing.

## 1 はじめに

ESG 投資は企業の環境 (Environment), 社会 (Social), ガバナンス (Governance) への取り組みを評価した投資である。近年, 企業の社会的責任 (CSR) が重要視され, 特に ESG は重要なキーワードである。

企業の ESG 文書開示数により ESG への積極性を評価した研究 [1] がある一方で, ESG というキーワードの曖昧さにより, 各企業の開示文書の内容から ESG への取り組みを評価することは難しい。文書を各トピックに割り当てる手法としてトピックモデルがある。トピックモデルの応用例として Word2Vec[2] などの単語の分散表現を用いたトピックモデルの構築 [3] や, ファンドの割り当てへの応用 [4] などがある。

そこで本研究では, ニューステキストの ESG トピックワードに着目した企業の ESG スコアの構築を行う。ESG のトピックワードの抽出には Word2Vec を使用し, 抽出した単語を用いてニューステキストの評価に使用する。また, その結果から ESG ファクターを構成し, これが既存の代表的なファクターに対して補完的に機能するかを検討する。

## 2 ESG 単語の抽出

### 2.1 単語の分散表現

単語の記号から類似度や関連性を計算するためには任意のベクトルで表現する必要がある, これを分散表現という。特に単語の分散表現を獲得する代表的な手法に Word2Vec[2] がある。Word2Vec はニューラル言語モデルであり, CBOW モデルと Skip-gram モデルの 2 つのモデルが提案されている。ネットワーク構造は Encoder と Decoder から構成され, Encoder は単語の記号から分散表現への変換, Decoder は分散表現から記号への変換を行う。Word2Vec から得られた単語の分散表現では, 図 1 のように似た単語同士は近いベクトルに表され, それらは一種のトピックの役割を果たす。したがって, 本研究ではその特性を利用して分散表現から「環境」, 「社会」, 「ガバナンス」の周辺単語を取り出すことで, ESG スコアの評価に使用する。

Word2Vec の問題として類義語だけでなく対義語も近いベクトルに変換されてしまう (例: 好き・嫌い, 高い・安い)。これは, ESG にネガティブなニュースに対しても高く評価しかねない。この問題を解決するために Word2Vec の学習においてはマーケットに対してポジティブなニュースのみを使用することでネガティブ単語を除去する。ポジティブなニュースを抽出する方法として QUICK ニュース解析サービス [5] が提供しているニューススコアを利用する。

\*連絡先: 茨城大学理工学科機械システム工学専攻  
〒316-8511 茨城県日立市中成沢町 4-12-1  
E-mail: 19nm401h@vc.ibaraki.ac.jp

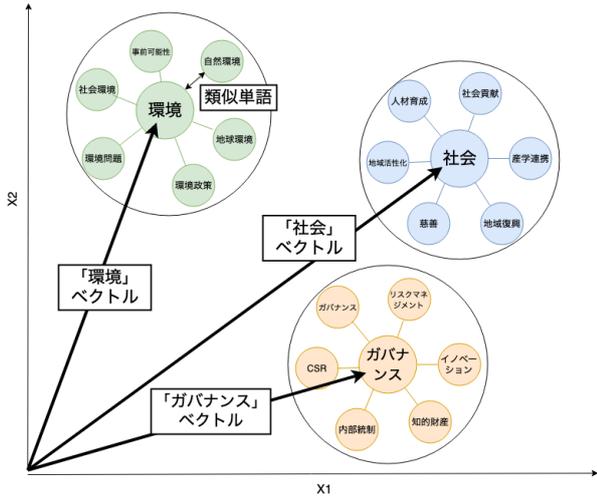


図 1: Word2Vec の概要

## 2.2 結果

Word2Vec の学習には 2011 年 1 月 1 日～2012 年 12 月 31 日の QUCIK ニュース解析サービスから配信されたニュースデータを使用する。Word2Vec による「環境」、「社会」、「ガバナンス」と  $\cos$  類似度が大きい上位 10 単語を表 1 に示す。

「環境」との  $\cos$  類似度が最も高い単語が「加速度」となった理由は、「加速度的増加」などのニュース特有の言い回しがあるためである。また、「社会」との  $\cos$  類似度が最も高い単語が「炭素」となった理由は、「低炭素社会」という単語がテキスト中に含まれるためである。

表 1: 各 ESG に類似する上位 10 単語

環境	社会	ガバナンス
加速度	炭素	委譲
永續	我が国	実効
リユース	CSR	総力
クラウドコンピューティング	事業	統制
保全	計算	戦略
親和	資本	選抜
国連	当期	疎通
低減	利益	法務
テレワーク	証券	一元化
カーボンフットプリント	評価	分業

## 3 ESG 評価

### 3.1 ESG スコア

Word2Vec を用いて「環境」、「社会」、「ガバナンス」に類似する上位 100 単語を取り出し、各企業の ESG スコアを計算する。例として「環境」スコアの計算方法を示す。ニュース  $i$  に含まれる Word2Vec から抽出した「環境」類似単語数を  $W_i$ 、「環境」単語のベクトルを  $w_E$ 、類似単語のベクトルを  $w_j$ 、類似単語を含むニュース数を  $N$  とすると、あるニュースの「環境」スコア  $Score_E$  を式 (1) のように定義する。

$$Score_E = \frac{1}{N} \sum_{i=1}^N \sum_{j=1}^{W_i} \cos(w_E, w_j) \quad (1)$$

ここで  $\cos(v_a, v_b)$  は単語  $a$  と  $b$  のベクトル  $v_a, v_b$  の  $\cos$  類似度を表す。

$$\cos(v_a, v_b) = \frac{v_a \cdot v_b}{|v_a| |v_b|} \quad (2)$$

なお、「社会」、「ガバナンス」も同様に計算する。

ESG は企業の長期的取り組みであることから、企業の ESG スコアは 2011 年 1 月 1 日から投資する月までの式 (1) の合計で計算し、スコアは月末に更新する。

### 3.2 ESG ポートフォリオ

上記で計算した企業の ESG スコアを上位 30% (High)、中位 40% (Medium)、下位 30% (Low) に分け、それぞれのパフォーマンスを計測する。なお、企業のニュース配信数によるサイズ効果をなくすため、ニュースが月に一度も配信されない銘柄はあらかじめ除去した。

期間は 2013 年 1 月 1 日～2018 年 12 月 31 日までの月足データを使用し、上記の ESG ポートフォリオについて累積収益率  $M_1(t)$  を計算する。

$$M_1(t) = M_1(1) + M_1(1) \prod_{\tau=1}^t \{1 + r(\tau)\} \quad (3)$$

$$r(t) = \frac{P(t) - P(t-1)}{P(t-1)} \quad (4)$$

ここで、 $P(t)$  は時刻  $t$  における株価、 $r(t)$  は収益率であり、初期資産は  $M_1(1) = 1$  とする。

得られた結果を図 2 に示す。ESG スコアが高い High が低い Low よりも高い収益率を得ており、Word2Vec で得られた ESG 単語が十分に機能していると示唆される。

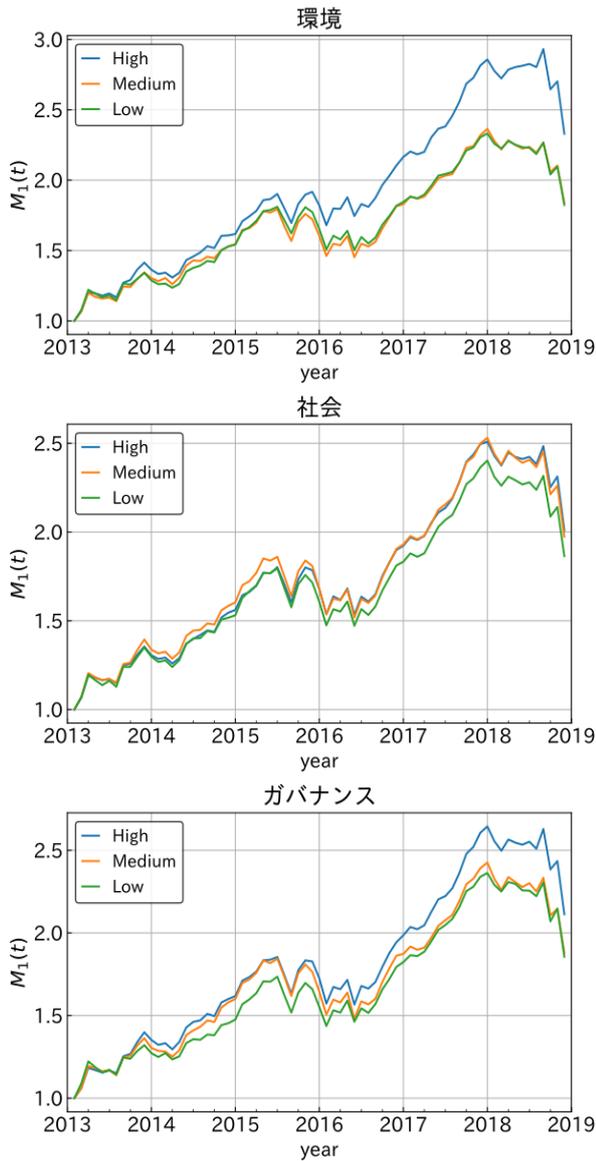


図 2: ESG ポートフォリオの累積収益率  $M_1(t)$

次に、High から Low のリターンを引くことにより、ESG ファクターとして累積収益率  $M_2(t)$  を計測した。

$$M_2(t) = M_2(1) + M_2(1) \prod_{\tau=1}^t \{1 + s(\tau)\} \quad (5)$$

$$s(t) = \frac{1}{N_H} \sum_{i \in High} r_i(t) - \frac{1}{N_L} \sum_{j \in Low} r_j(t) \quad (6)$$

$r_i(t)$  は時刻  $t$  におけるポートフォリオ  $i$  の収益率、 $N_H$  は上位ポートフォリオの銘柄数、 $N_L$  は下位ポートフォリオの銘柄数、 $s(t)$  は各 ESG ファクターのリターンである。初期資産は  $M_2(1) = 1$  とする、

ここで、得られた  $s(t)$  の平均、標準偏差、平均/標準偏差を表 3 に、累積収益率  $M_2(t)$  を図 3 に示す。これ

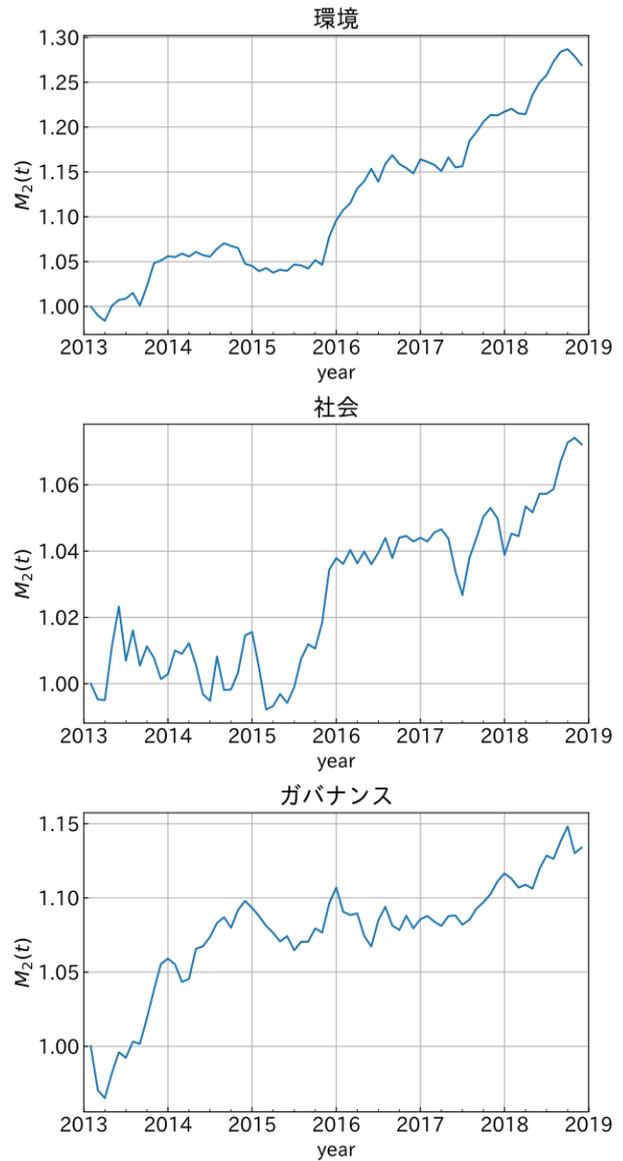


図 3: ESG ファクターの累積収益率  $M_2(t)$

らの結果から、ESG の各ファクターにおいて正の収益を期待でき、特に「環境」と「ガバナンス」に関して安定的なリターンを確認できる。

### 3.3 マルチファクターモデル

前章で得られた各企業の ESG スコアの有効性を 3 ファクターモデル [6] により検証する。3 ファクターモデルは次式の通りである。

$$r_{it} - r_{ft} = \beta_i + \beta_i^M (r_{Mt} - r_{ft}) + \beta_i^{SMB} SMB_t + \beta_i^{HML} HML_t + \epsilon_{it} \quad (7)$$

表 2: ファクターの相関係数

	マーケット	サイズ	バリュウ	環境	社会	ガバナンス
マーケット	1.00	-0.01	-0.14	-0.18	-0.38	-0.15
サイズ	-0.01	1.00	-0.46	-0.05	-0.14	-0.31
バリュウ	-0.14	-0.46	1.00	-0.23	0.21	0.36
環境	-0.18	-0.05	-0.23	1.00	0.29	0.37
社会	-0.38	-0.14	0.22	0.29	1.00	0.32
ガバナンス	-0.15	-0.31	0.36	0.37	0.32	1.00

表 3: ESG ファクターのリターン

	平均 (%)	標準偏差 (%)	平均/標準偏差
環境	0.31	0.98	0.32
社会	0.09	0.67	0.13
ガバナンス	0.18	0.92	0.20

表 4: ESG ファクターの 3 ファクターモデルによる  
偏回帰係数

	環境		社会		ガバナンス	
	係数	t 値	係数	t 値	係数	t 値
$\beta_i$	0.36	3.16***	0.15	1.91	0.24	2.32*
$\beta_i^M$	-0.06	-1.99*	-0.07	-3.24***	-0.03	-1.01
$\beta_i^{SMB}$	-0.10	-1.75	-0.03	-0.73	-0.08	-1.54
$\beta_i^{HML}$	-0.14	-2.81***	0.03	0.96	0.09	2.03*

\* $p < 0.05$ , \*\*\* $p < 0.001$

ここで、 $r_{Mt} - r_{it}$  はマーケットファクター、 $SMB_t$  はサイズファクター、 $HML_t$  はバリュウファクターであり、 $\beta_i^M, \beta_i^{SMB}, \beta_i^{HML}$  はそれぞれの偏回帰係数を表す。

ESG ファクターの有効性を検証するために、式 (7) の  $r_{it}$  を ESG ファクターのリターン  $s(t)$  として重回帰分析を行う。この分析結果を表 4 に示す。「環境」と「ガバナンス」に関しては、切片が有意に正となり、3 ファクターでは説明できないリターンが存在することが示唆される。また、3 ファクターと  $s(t)$  との相関係数を表 2 に示す。この結果として、作成した ESG ファクターは他のファクターとの相関が小さいため、独自のファクターとして機能していることが分かる。

## 4 まとめ

本研究では Word2Vec によりニューステキストから抽出した ESG トピックワードを用いて各企業の ESG スコアを評価し、その評価に基づいて構成した ESG ファクターを分析した。その結果、「環境」および「ガ

バナンス」ファクターが既存の代表的なファクターでは説明できないリターンを有することが確認できた。

今後の課題として、「社会」ファクターのパフォーマンスの向上や単語だけでなく文書構造を考慮した ESG スコアの評価などが挙げられる。

なお本研究の一部は、文化省科研費基礎研究 (C) (No.16K00320) の助成により行われました。

## 参考文献

- [1] 湯山智教, 白須洋子, 森平総一郎. ESG 開示スコアとパフォーマンス, 証券アナリストジャーナル, Vol. 57, No. 10, pp. 72-83, 2019
- [2] Tomas Mikolov, Kai Chen, Greg Corrado, and Jeffrey Dean. Efficient Estimation of Word Representations in Vector Space, *International Conference on Learning Representations*, pp. 1-12, 2013.
- [3] Rajarshi Das, Manzil Zaheer, and Chris Dyer. Gaussian LDA for Topic Models with Word Embeddings, *In Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics*, pp. 795-804, 2015.
- [4] Masanori Hirano, Hiroki Sakaji, Shoko Kimura, Kiyoshi Izumi, Hiroyasu Matsushima, Shintaro Nagao, and Atsuo Kato. Related Stocks Selection with Data Collaboration Using Text Mining, *Information*, Vol. 10, No. 3, pp. 1-17, 2019
- [5] 株式会社 QUICK, QUICK ニュース解析サービス, <https://nwfs.newsdolphin.jp/ja>, (参照 2019-02-23)
- [6] Eugen F.Fama and Kenneth R. French. Common Risk Factors in the Returns on Stocks and Bonds, *Journal of Financial Economics*, Vol. 33, No. 1, pp. 3-56, 1993