

ホライゾン・スキヤニングの自動化のための言語処理応用

石垣 達也¹ 西野 涼子^{1,2} 鷲野 壮平¹ 五十嵐 広希³ 村井 昭彦¹ 永井 由佳里² 鷲田 祐一⁴

¹ 産業技術総合研究所 ² 北陸先端科学技術大学院大学 ³ 東京大学 ⁴ 一橋大学

{ishigaki.tatsuya, s.washino, hk-igarashi, a.murai}@aist.go.jp

{suzuko.k, ynagai}@jaist.ac.jp b101348r@r.hit-u.ac.jp

概要

本稿では未来学分野におけるホライゾン・スキヤニング [1] の自動化を目指し、文書探索、コメント生成問題およびデータセットを提案する。ホライゾン・スキヤニングは未来に起こりうる社会変化を分析する枠組みで、1) 新聞記事等を大量に収集し、2) 重要な記事には主観的な意見や記事の要約をコメントとして記述する。収集された記事とコメントは専門家による社会変化についてのシナリオ記述に活用される。本研究ではデータセットの分析、探索およびコメント生成モデルを構築し自動評価および未来学の専門家が実際に使用することで評価し、言語処理技術の応用可能性を検証する。

1 はじめに

未来学の分野では、中長期的な未来に起こりうる大きな社会変化やリスクについて、シナリオを記述する取り組みが行われている。例えば、「自動消毒ロボットは将来、駅だけでなくオフィスビルや飲食店など他の業種にも幅広く適用され、人手不足の深刻な人間の清掃員の代替となるだろう」というシナリオは、企業や政府における意思決定に役立てられる。高品質なシナリオを生成するための手法について、当該分野では活発に研究 [2] されているが、その多くは専門家による情報収集に依存しておりコストが問題となる。言語処理技術の応用はこのような課題の一助になりうる。

本研究では代表的なシナリオ作成手法である Systemic Foresight Methodology (SFM) [3] に着目する。この手法は、シナリオ作成を 1) ホライゾン・スキヤニング、2) プランニングという 2 つの部分問題に分割して解く。前者は情報収集を行うステップで、社会変化の兆候を示唆する情報を含む新聞記事等を大量に収集し、各記事にコメントを付与する。後者はシナリオを記述するステップで、収集した記事を専門家同士で議論しながら集約し最終的に数個のシナリオにまとめ

る。本稿では後段の処理に影響を与えることからより重要なホライゾン・スキヤニングに着目し、1) 未来を示唆する文書の探索問題、2) 主観的な意見を含むコメント生成という自然言語処理における新たな問題設定と、3) これらのためのデータセットを提案する。

図 1 に現在専門家により手作業で行われている、ホライゾン・スキヤニングの手順 [4, 1] を示す。入力は大規模なテキスト集合で、本研究では一例として日本語の新聞記事を想定する。出力は事前に定義されたフォーマットによる文書で、スキヤニングマテリアルと呼ばれる。ホライゾン・スキヤニングでは、まず入力記事集合から未来の社会変化を示唆する記事を探る。例えば、図 1 の記事 2 のような「高輪ゲートウェイ駅での消毒ロボットの実証実験」についての記事は、鉄道以外の産業にも将来的にロボットが配置されるという社会変化を示唆するため、探索されるべき記事である。一方、記事 4 は天気予報であり、未来についての言及を含むが日常的に必要な情報を提示しているに過ぎず、中長期的な社会変化について示唆しない。よって、探索されるべきではない。スキヤニングマテリアルは未来の社会変化を示唆する各記事に対し作成され、記事タイトル、要約、主観的なコメントを含む。コメント部分には記事の短い要約、主観的な意見、主張や将来起こりうる社会変化などが記述される。探索した新聞記事本文を入力し、主観的なコメントを自動生成する問題を 2 つ目の設定として扱う。本研究では特にスキヤニングマテリアル作成のための記事の探索とコメント作成に着目し、記事タイトルと要約部分については要約技術等の既存技術を用いる想定ととし、研究対象としない。

本研究で提案する文書探索問題は、将来の社会の変化を示唆するか否かを考慮する点において、既存の文書探索の設定 (例えば [5] など) と異なる。コメント生成問題は言語生成問題の一つとみなすことができるが、将来起こりうるイベントについて記述する点に特徴がある。

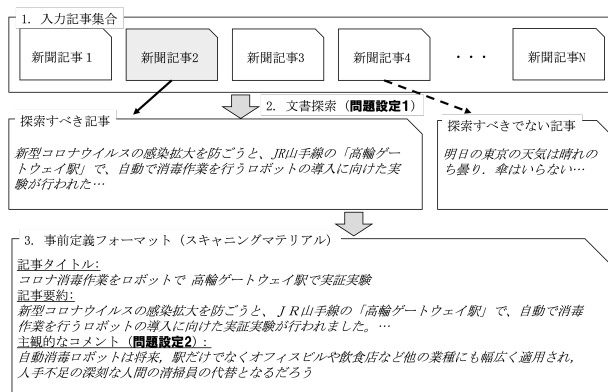


図1 ホライゾン・スキャンングの手順.

2,266 の未来の社会変化を示唆する新聞記事を SFM の知識のある者が収集し、データセットを作成した。分析より、これらの一般的な新聞記事とは意味的な距離が離れており、異なる特徴語を持つことがわかった。このデータセットから学習した文書探索モデルは、BERT による手法は適合率@100 において 70% の性能を示し、専門家による実際のシナリオ作成の実用において、十分耐えうることを人手評価により確認した。コメント生成問題では、要約タスク等において良い性能を示す BART [6] であっても、主観的なコメントを生成するのが難しく挑戦的な課題であることが分かった。

2 データセット作成

データセットの作成には SFM に関する専門知識が必要となる。したがって、本研究では専門知識を持つ者が未来の社会変化を示唆する記事を収集することが望ましい。そこで、ホライゾン・スキャンングに関する講義を担当する大学教員とその講義を受講し専門知識を持つ学生らが、2003 年から 2020 年の 17 年間に渡り講義の一環として収集した記事の提供を受け、データセットとした。合計 33 人が以下の指示の下、記事収集とコメント付与を行った: 1) 中長期的な未来に起こりうる社会変化を示唆する多様な記事を Web 上から手作業で探索する、2) タイトル、要約、コメントを記述する。コメントには収集者の主観的な意見や将来に起こりうる中長期的な変化を記述する、3) 最後に図 1 に例示するスキャンングマテリアルのフォーマットにまとめる。最終的に 2,266 記事を収集し、各記事に平均 2.8 文のコメントが人手で付与された。

3 分析

3.1 一般的な新聞記事との意味的な距離

探索されるべき記事は一般的な記事とは意味的な距離が離れており、異なる特徴を持つことを定量的に確かめる。2020 年の朝日新聞コーパス¹⁾中の記事を一般的な記事とし、各記事中の名詞に対する日本語 word2vec [7]²⁾の平均ベクトルを記事を表現する分散表現とする。すべての記事に対し分散表現を計算し、その中心ベクトル \mathbf{c}_g を一般的な新聞記事を表現したベクトルとみなす。 \mathbf{c}_g から提案データセット中の各記事を表現する分散表現 \mathbf{d} への距離を、以下のようにコサイン類似度を用いて定義する: $\text{dist}(\mathbf{c}_g, \mathbf{d}) = 1 - \text{cosine}(\mathbf{c}_g, \mathbf{d})$ 。この距離の提案データセット中のすべての記事での平均は 0.34 であった。一方、中心ベクトルから朝日新聞コーパス中の記事への距離は平均すると 0.30 であった。この差が統計的に有意であることを t 検定を用いて確認した。よって、一般的な新聞記事と意味空間上での距離が離れている特徴が確かめられた。

3.2 特徴語

探索されるべき記事の特徴語を使用語彙の観点から分析する。具体的には次の式のように単語出現頻度 (TF) の差を用い、単語 x が探索されるべき記事をどの程度特徴づけるかを表現するスコアを計算する: $\text{diff}(x) = \text{TF}_C(x) \text{IDF}_{C+G}(x) - \text{TF}_G(x) \text{IDF}_{C+G}(x)$ 。ここで、 $\text{TF}_C(x)$ と $\text{TF}_G(x)$ はそれぞれ、探索されるべき記事の集合 C と一般的な記事集合 G での x の出現頻度である。 IDF_{C+G} は C と G の和集合における IDF で、機能語などの多くの文書に出現する一般的な語のスコアを抑制する。 $\text{diff}(x)$ が正の方向に振れると x は探索対象とする記事の特徴語であり、 $\text{diff}(x)$ が負の方向に振れると一般的な記事の特徴語である。

結果、正の方向に振れた語には「ツイッターの新機能」「台湾での干ばつ」といったように、稀に起こるイベントについての言及はそれぞれ 22.31, 20.88 と高いスコアを得た。また、「テクノロジー」のスコアは 2.34 であり、技術分野に関する語も高いスコア得た。一方、一般的な新聞記事の特徴づける語としては「知事の会見」「容疑者の逮捕」など、日常的に起こりう

1) <http://www.asahi.com/information/cd/gakujutsu.html>

2) 東北大の配布する日本語 Wikipedia から学習した単語ベクトルを用いる。 http://www.cl.ecei.tohoku.ac.jp/~m-suzuki/jawiki_vector/

表 1 意味ラベルの定義と例.

入力記事	
トミーは、バイオリンの音が出るおもちゃ「Evio（エビオ）」を7月末に発売する。バイオリニストのように音楽を奏でることができる。価格は7,000円。小学生の女の子を中心に、幅広い年齢層からの需要を見込んでいる。	
意味ラベル	定義と例文
要約	入力記事の要約 「トミーがバイオリンの音を奏でるおもちゃを発売する。」
背景	記事に含まれない客観的事実 「様々なおもちゃが近年発売されている。」
主張(現在)	事実に対する意見 「ITとおもちゃの融合はとても良いことだ。」
主張(未来)	未来の変化についての意見 「画期的なおもちゃ作りにITは不可欠となるはずだ。」
見込み	将来起こりうるイベント 「おもちゃだけでなくカラオケも進化するのかもしれない。」
感想	意見を含まない感想 「面白いおもちゃだと思う。」

るイベントについての言及は-0.89, -0.90と負の方向に振れ、一般的な記事に特徴的な語と認識された。より詳細な単語-スコア対の一覧は付録に含める。

3.3 意味ラベル付与によるコメント分析

次にコメント部分についての分析を行う。提案データセットから無作為に30コメントを抽出し、手作業でコメントに含まれる各文に対し表1に定義する、意味ラベルを付与する。付与作業においては1つの文に対し、複数ラベルの付与を許容する。例えば、「トミーがバイオリンのおもちゃを発売した」という記事に対し、「様々なおもちゃが近年発売されている。」というコメント記述は“背景”とラベル付ける。

結果、83%のコメント(30のうち25)は“要約”もしくは“背景”という事実について言及する文から始まり、“主張”や“見込み”といった主観的な意見へと続いた。30コメントのうち4コメントは事実を含まず、主観的な意見のみ記述された。ただ1つ例外的に主観的な意見を含まず、“感想”のみのコメントが存在した。客観的な記述と主張などの主観的な記述を含む83%部分である25コメントに着目すると、うち76%(25のうち19)は“要約”から始まり、24%(25のうち6)は元の記事には含まれない“背景”情報を提示した。

4 モデル

文書探索モデルは記事集合を入力とし、各記事に対し中長期的な社会変化を示唆するか否かを表現するスコアを計算し、スコアが高い順に並べ替え出力する。ス

表 2 文書探索モデルの適合率, 再現率, F 値.

	適合率	再現率	F 値
BERT	58.6	57.8	58.1
意味的な距離	5.71	65.6	10.5
IDF	8.60	33.4	13.7

表 3 文書探索モデルのランキング性能.

P@N	5	10	30	50	100
BERT	80.0	80.0	76.7	76.0	70.0
意味的な距離	40.0	20.0	6.7	4.0	3.0
IDF	40.0	20.0	16.7	16.0	15.0

コア計算の手法として、1) BERT [8] による二値分類、2) 一般的な新聞記事からの距離および 3) IDF を用いたヒューリスティックを比較する。BERT による手法は提案データセット中の記事を正例、朝日新聞コーパスからサンプルした記事を負例として二値分類器を学習する。一般的な記事からの距離は朝日新聞コーパスからサンプルした記事集合の中心ベクトル \mathbf{c}_g とスコアリング対象記事の分散表現 \mathbf{x} を用い、スコアを $1 - \cos(\mathbf{c}_g, \mathbf{x})$ として計算する。ヒューリスティックによる手法は、探索されるべき記事には特異な語が多いと考え、スコアリング対象記事中の名詞の IDF の平均値をスコアに用いる: $\frac{\sum_i \text{IDF}(x_i)}{n}$ 。

コメント生成モデルは探索した記事を1つ入力し、コメントを出力する。本研究では単一文書要約において良い性能が報告されている BART [6] を提案データセットコメントを用いてファインチューニングする。

5 実験と評価

5.1 データとパラメータ

BERT の学習には提案データセットのうち 2020 年に発行された 320 記事を正例として用い、2020 年に発行された朝日新聞記事のうち 8000 件を無作為に抽出し負例とする。データセットを 3:1:1 の割合に分割し、それぞれ学習、開発、評価セットとして用いる。二値分類器の出力スコアがしきい値 δ 以上の場合に分類対象記事を正例であると判定し、 δ は開発セットを用いて決定する。BART には京大の配布するモデル³⁾を用いる。自動要約においてよく用いられるパラメータ設定⁴⁾でファインチューニングする。

表 4 被験者が作成したシナリオについての統計.

	シナリオ数
被験者により作成されたシナリオ	85
自動と人手両方を基にしたシナリオ	71
自動探索記事を基にしたシナリオ	11
人手収集記事を基にしたシナリオ	3

5.2 文書探索の自動評価

表 2 に探索モデルの適合率, 再現率, F 値を示す. また, 表 3 に適合率@N を N の値を 5 から 100 まで変化した値を示す. BERT による手法は F 値において 58.1 で, 意味的な距離による手法の 10.5 や IDF の 13.7 よりも良い. 適合率@N においても, N=5 の設定で BERT による手法は 0.80 を示し, 意味的な距離や IDF による手法の 0.40 よりも高い値を示した. N の値を大きくすると, 教師あり学習に基づく手法は N = 100 まで大きくしても値の低下は 0.70 に抑えられるのに対し, 意味的な距離に基づく手法および IDF による手法はそれぞれ 0.02 と 0.09 で, 性能が著しく低下する. 提案データセットによる教師あり学習は, 適合率@N の観点からもより頑健なモデルといえる.

5.3 自動探索文書の実利用による人手評価

既存評価指標だけでなく, 自動探索記事を専門家によるシナリオ作成に実際に用いる実験を行い, 実利用での効果を検証する. この実験では, まず, BERT が高いスコアを与えた 75 記事と人手収集した 75 記事を実験的に混ぜ合わせ, 合計 150 記事からなる記事集合を作成する. SFM の知識を持つ 17 人に 150 記事を提示し, 記事集合を目視しながら一人あたり 5 シナリオを作成させ, 合計 85 (17 人 × 5) のシナリオを得る. 各シナリオは 1 つ以上の記事を基に作成され, シナリオ作成の際に参考にした記事も併せてアノテーションさせた. 結果, 1 つのシナリオは平均して 4.8 個の記事を基に作成された. 表 4 に被験者が作成したシナリオの統計を示す. 85 シナリオのうち 71 シナリオは自動探索した記事と人手収集した記事を組み合わせで作成したものであった. 探索文書のみを組み合わせで作成したシナリオが 11 存在し, これは人手収集記事のみを組み合わせで作成した 3 つのシナリオよりも多い. これらを合計すると, 85 シナリオのうち 96.4% (82 シナリオ) という高い割合で, 1 記事以上の自動探索記事が専門家たちに活用されたことが確かめられた.

- 3) https://github.com/utanaka2000/fairseq/blob/japanese_bart_pretrained_model/JAPANESE_BART_README.md
- 4) <https://github.com/pytorch/fairseq/blob/main/examples/bart/README.summarization.md>

表 5 コメント生成問題の自動評価結果.

	ROUGE-1	ROUGE-2	ROUGE-L
BART	34.85	7.52	21.58
リード 3	30.05	8.15	19.07

5.4 コメント生成の自動評価

表 5 に ROUGE の値を示す. BART は ROUGE-1 と ROUGE-L においてリード 3 よりも高い値を示した. 一方, ROUGE-2 においてはリード 3 法よりも低い値であった. 3.3 節で述べたように人間のコメントの冒頭部分には入力文書の要約が含まれやすく, この要約部分の多くは原文書を言い換えない抽出型の要約となっている. BART が原文書の言い換えを許容する生成型の要約手法であるのに対し, リード 3 は文抽出型の手法である. よって, 正解コメントとリード 3 法の出力で単語の重複が多くなり, ROUGE-2 の値が大きくなる. 実際の生成コメントは付録に例示する.

5.5 コメントの採点による人手評価

人手採点による評価結果を付録の表 6 に示す. 表に示す値は行に示す手法が列に示す手法より良いと判定された回数である. 上から順に 1) 未来の変化を想起させるか, 2) 流暢さ, 3) 正しさの 3 つの観点における値を示す. 結果, 未来の変化を想起させるかという観点では, BART が 51 回リード 3 よりも良いと判定されたのに対し, リード 3 は BART よりも良いと判定されたのは 41 回であった. よって, BART の出力にはリード 3 と比較し, より未来の変化を想起させる記述が含まれる. 一方, 参照コメントは BART よりも 72 回良いと判定されており, BART がより良いと判定された回数はわずか 18 回である. 付録の出力例に示すように, 要約などの客観的な記述と意見などの主観的な記述を混在させた出力を得るのが難しい. 言語生成モデルには言語の客観性, 主観性の制御といった観点からの拡張が必要である.

6 おわりに

ホライゾン・スキニングを自動化するため, 文書探索及び言語生成とデータセットを提案した. BERT による手法は自動および人手評価により, 実用的な性能が得られることを確認した. コメント生成は, 最先端の言語生成モデルを用いたとしても主観的なコメントを正しく生成するのが難しく挑戦的な課題であることが分かった. 主観的な記述と客観的な記述を正しく制御する新たな言語生成技術が必要であろう.

謝辞

本研究の成果は LREC2022 [9] および KICSS2022(Nishino, Washida, Ishigaki, Washino, Igarashi, Murai, and Nagai 2022) の発表論文を基にした。本研究は、JST 次世代研究者挑戦的研究プログラム JPMJSP2102 の支援を受けたものです。整備したデータセットの元となる記事およびコメントを提供くださった一橋大学の MBA コース学生の皆様、ワークショップの参加者に感謝します。人手評価におけるデータ収集は、JAIST の倫理審査委員会の承認を得ています。計算には産総研の AI 橋渡しクラウド (ABCI) を利用しました。実験には朝日新聞コーパスを使用しました。

参考文献

- [1] Ziauddin Sardar. The namesake: Futures; futures studies; futurology; futuristic; foresight—what’s in a name? *Futures*, Vol. 42, No. 3, pp. 177–184, 2010.
- [2] Yuichi Washida and Akihisa Yahata. Predictive value of horizon scanning for future scenarios. *foresight*, 2020.
- [3] Ozcan Saritas. Systemic foresight methodology. In *Science, technology and innovation policy for the future*, pp. 83–117. Springer, 2013.
- [4] Yuichi Washida and Akihisa Yahata. Predictive value of horizon scanning for future scenarios. *Foresight*, Vol. 23, No. 1, pp. 17–32, 2021.
- [5] Bhaskar Mitra, Nick Craswell, et al. An introduction to neural information retrieval. Now Foundations and Trends, 2018.
- [6] Mike Lewis, Yinhan Liu, Naman Goyal, Marjan Ghazvininejad, Abdelrahman Mohamed, Omer Levy, Veselin Stoyanov, and Luke Zettlemoyer. BART: Denoising sequence-to-sequence pre-training for natural language generation, translation, and comprehension. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics (ACL2020)*, pp. 7871–7880, Online, 2020. Association for Computational Linguistics.
- [7] Tomas Mikolov, Ilya Sutskever, Kai Chen, Greg S Corrado, and Jeff Dean. Distributed representations of words and phrases and their compositionality. In C. J. C. Burges, L. Bottou, M. Welling, Z. Ghahramani, and K. Q. Weinberger, editors, *Advances in Neural Information Processing Systems*, Vol. 26. Curran Associates, Inc., 2013.
- [8] Jacob Devlin, Ming-Wei Chang, Kenton Lee, 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 (NAACL2019)*, pp. 4171–4186, Minnesota, USA, 2019. Association for Computational Linguistics.
- [9] Tatsuya Ishigaki, Suzuko Nishino, Sohei Washino, Hiroki Igarashi, Yukari Nagai, Yuichi Washida, and Akihiko Mu-

rai. Automating horizon scanning in future studies. In *Proceedings of the Thirteenth Language Resources and Evaluation Conference*, pp. 319–327, Marseille, France, June 2022. European Language Resources Association.

付録

	参照コメント	BART	リード3
未来の示唆			
参照コメント	-	72	77
BART	29	-	51
リード3	29	41	-
流暢さ			
参照コメント	-	69	37
BART	29	-	33
リード3	46	65	-
正しさ			
参照コメント	-	49	16
BART	36	-	18
リード3	92	90	-

表 6 人間による採点評価の結果。人間によるコメント、BART およびリード3を比較し、行の手法が列の手法よりも良いと判定された回数を示す。

1. 入力記事:
JR 東日本は、高輪ゲートウェイ駅において、自動移動式消毒ロボットの実証実験を開始すると発表した。夜間にロボットを稼働させ、効果を検証する。導入されるロボットは3台（中略）。
人間:
鉄道会社では、事業を維持するために必要な業務について、人材を確保することができない。自動改札機や自動精算機などはその代表的な製品ですが、将来的には社会のさまざまな場面で応用できるようになると考えられる。
BART:
スプレーを搭載した自動移動式消毒ロボットが登場した。本技術は、外食産業における消毒・洗浄に応用できる可能性がある。
2. 入力記事:
徳島大学発のベンチャー企業グリラスは、食用コオロギ事業を拡大するため、総額2億3000万円を調達した。今回の資金調達を機に、株式公開に向けた体制整備を進める予定だ。グリラスは5月にジェイテクトと業務提携し、全自動生産システムを共同開発した。
人間:
最近、日本でも10匹と30匹のコオロギを使ったフィナンシェが発売され、2日で売り切れたそう。私はおせんべい以外、コオロギを使った食品を見たことがないが、いよいよ昆虫食が社会に浸透していく兆しがあるようだ。
BART:
徳島大学発のベンチャー企業グリラスが、食用コオロギの事業拡大のため、ベンチャーキャピタルから総額2億3000万円の資金調達を行った。全自動生産システムの共同開発に取り組んでいる。

表 7 コメントの生成例。