

2

AI と IoT

## 2-2 単語分散表現を用いた時系列データによる AI ブーム分析 AI Boom Analysis Using Time Series Data in Word Embeddings

徳島大学

谷岡広樹

Tokushima University

Hiroki Tanioka

### 【趣旨】

人工知能（Artificial Intelligence：AI）の技術発展は、現在の人間の「しごと」の多くを代替することで、産業構造に変革をもたらすとされている。しかし、この変革により新たな「しごと」が生み出されるであろうことは過去の産業革命における「しごと」の変化からも明らかであろう。本稿では、単語分散表現（Word Embeddings）を用いて新聞記事等から人工知能に関連するキーワードの時系列データを抽出することで、第四次産業革命のキーファクターの一つである人工知能が、どのように日本社会で認知され注目されるようになってきたか、その変化を明らかにする。

### 【キーワード】

人工知能、単語分散表現、時系列データ、テキストマイニング

Replacing many of "work" of humanity with Artificial Intelligence (AI) is said to bring a big change in the industrial structure. However, this change will create a new "work". It will be evident from the change of "work" in the past industrial revolution. This paper describes how artificial intelligence, one of the key factors of the Fourth Industrial Revolution, has become recognized and noticed in Japanese society. We clarify the changes of the time series by investigating the related keywords of artificial intelligence using word embeddings on newspaper articles, etc.

### 【Keywords】

Artificial Intelligence, Word Embeddings, Time Series Data, Text Mining

## 1 ◆ はじめに

人工知能の技術の発達で、将来、我々人間の「しごと」を奪うかもしれないという議論が巻き起こり、「人工知能」や「AI」といったものがいったい何者かについて、若年層から高齢者まで幅広い年齢層で関心を広めている。著者に対しても、徳島県内の高等学校から「AIとはなにか？」をテーマとした出張授業<sup>1</sup>の依頼が増加している。受講生からは、「AIが私たちのしごとを奪う

1 2018年度は徳島市立高校、徳島県立城北高校、徳島県立徳島北高校の3校、2019年度は徳島市立高校、徳島県立徳島文理高校、徳島県立城北高校の3校

といわれているが、実際にそのようになるのか？先生はどう考えるのか？」といったような質問も度々受けるようになった。その回答としては、「現在の AI は人間と同等の判断ができるものではなく、人間の知能の一部を代替する弱い AI と呼ばれているものなので、今のところ心配することはない。むしろ賢い AI を作ったり、うまく使いこなしたりできるようになってほしい。」といったように著者自身の持論を踏まえて説明している。このように、「人工知能」や「AI」は現在の日本社会および日本経済において空前のブームであるといえる。そこで、この AI ブームがいつ、どの程度、どのようなカテゴリにおいて沸き起こったのかについて分析する。具体的には、「人工知能」に関するトピックについて、深層学習を用いて抽出された関連キーワードについて、複数のメディアを対象とした時系列データ上で出現頻度を計算し、AI ブームの盛り上がりの状況を客観的にみる。

## 2◆背景


### 2.1. 人工知能の発達

人工知能（Artificial Intelligence）の技術発展は、就業構造の変革に深く関わっている。人工知能が経理や給与計算等のバックオフィスの「しごと」を代替できるようになることは勿論、人工知能を搭載した IoT 機器やロボットの出現により、工場の製造ラインや商品の販売員、コールセンター等、様々な場面で人間と代替可能であると予想される。一方で、人工知能を作る「しごと」や、人工知能の出力結果をみて判断しながら共に働く「しごと」の一部には、これからも人間の活躍の場が存在すると考えられるが、就業構造変革の成否が産業構造に影響を与え、人間に求められる「しごと」の中身が変わろうとしていることは間違いない。

### 2.2. 第四次産業革命

モノとインターネットがつながる Internet of Things (IoT) や AI といった技術によって起こされる第四次産業革命は、ドイツで提唱された Industry 4.0 やアメリカの Industrial Internet といった技術革新に関する取り組みを背景としており、これらの取り組みによって大きな産業・技術革新が起きる可能性が示唆されている。この革新により現状のまま放置した場合と対比する形で目指すべき「しごと」の姿を表 2.1 に示す [経済産業省 2016]。本稿では、日本経済が向かうと思われる就業構造改革のキーファクターの一つである“人工知能”に着目して、そのトレンドを分析する。

表 2.1 第4次産業革命による就業構造変革の姿（[経済産業省 2016] より抜粋・修正）

現状放置 市場喪失し、仕事の量は減り、質も低下	就業構造変革 	目指すべき姿 グローバル市場を獲得し、質・量ともに十分な仕事
海外に流出	AI やロボット等を創り、新たなビジネスのトレンドを創出する仕事 (例) ・グローバル企業の経営戦略策定、トップレベルのデータサイエンティスト・研究開発 等	内外から集積
大きく減少 低付加価値な製造ラインの工員・営業販売・バックオフィス等	AI やロボット等を使って、共に働く仕事 (例) ・様々なビジネスの企画立案 ・データサイエンティスト等のハイスキルの仕事のサポート業務(ビジネスプロセスの変化をオペレーションレベルに落とし込む橋渡し役) ・今後激増するカスタマイズ化された商品・サービスの企画・マーケティング	新たな雇用ニーズに対応
	AI やロボット等と住み分けた仕事 (例) ヒューマン・インタラクション ・人が直接対応することがサービスの質・価値の向上につながる高付加価値な営業・販売やサービス	
多くの仕事が低賃金化	AI やロボット等に代替される仕事 (例) ・低額・定型の保険商品の販売員、スーパーのレジ係等（高代替確率） ・大衆飲食店の店員、コールセンター 等（高代替確率） ・経理、給与管理等の人事部門、データ入力係 等（変革の成否を問わず減少） ・建設作業員 等	AI やロボットで代替

### 3◆人工知能とテキストマイニング

一言で人工知能といっても、実際には様々な技術、様々な応用範囲が考えられる。本稿では、人工知能の一分野である自然言語処理技術から、深層学習を用いた word2vec を採用し、単語分散表現を用いたテキストマイニングを行う方法とその結果について述べる。

#### 3.1. 単語分散表現

単語分散表現とは、トピックモデルと呼ばれる手法でキーワードの特徴を分析し、その特徴をベクトルデータで表現したものである。LSA (Latent Semantic Analysis) (Deerwester, et al. 1990)、LDA (Latent Dirichlet Allocation) (Blei, Ng and Jordan 2003)、word2vec (Mikolov, et al. 2013) 等を用いたものがあるが、日本語を対象とした場合は、まず形態素解析器等を用いて文章を分かち書きし、適切な単位の単語（キーワード）を得てから、文や文書単位でキーワード間の関係性を分析することで得られる。

### 3.1.1. 形態素解析

形態素解析器としてはSudachi (Takaoka, et al. 2018) を用いる。Sudachi は形態素解析辞書として、国立国語研究所が公開している日本語辞書 UniDic (伝、ほか 2007) を利用しており、比較的新しい単語を適切な区切りで得られると考えた。

### 3.1.2. word2vec

深層学習 (Deep Learning) を用いて単語分散表現を得る手法はいくつか提案されているが、本研究では、国語研日本語ウェブコーパス NWJC<sup>2</sup> を Sudachi で解析し、word2vec で学習した学習済みモデル (nwjc 20190314 版、語数 3,644,628、次元数 300) (真鍋、ほか 2018) を用いる。

## 3.2. テキストマイニング

テキストマイニング (Text Mining) は、文字列を対象としたデータマイニングのことである。文章を文、文節、単語等で区切り、それらの出現頻度や共起頻度や相関、時系列の傾向等を解析するための手法を指す。

「人工知能 (AI) の進化を受け、データの活用法は企業の競争力を左右する。」



(形態素解析)

人工知能	名詞, 普通名詞, 一般, *, *, *	人工知能	人工知能	ジニコウチノウ	0
(	補助記号, 括弧開, *, *, *, *	(	(	キゴウ	0
AI	名詞, 普通名詞, 一般, *, *, *AI	AI	エーアイ	0	
)	補助記号, 括弧閉, *, *, *, *	)	キゴウ	0	
の	助詞, 格助詞, *, *, *, *	の	ノ	0	
進化	名詞, 普通名詞, サ変可能, *, *, *	進化	進化	シンカ	0
を	助詞, 格助詞, *, *, *, *	を	ヲ	0	
受け	動詞, 一般, *, *, 下一段-カ行, 連用形-一般	受ける	受ける	ウケ	0
,	補助記号, 読点, *, *, *, *	,	,	0	
データ	名詞, 普通名詞, 一般, *, *, *データ	データ	データ	0	
の	助詞, 格助詞, *, *, *, *	の	ノ	0	
活用	名詞, 普通名詞, サ変可能, *, *, *	活用	活用	カツヨウ	0
法	名詞, 普通名詞, 一般, *, *, *法	法	ハウ	0	
は	助詞, 係助詞, *, *, *, *	は	ハ	0	
企業	名詞, 普通名詞, 一般, *, *, *企業	企業	キギョウ	0	
の	助詞, 格助詞, *, *, *, *	の	ノ	0	
競争力	名詞, 普通名詞, 一般, *, *, *競争力	競争力	キョウソウリョク	0	
を	助詞, 格助詞, *, *, *, *	を	ヲ	0	
左右	名詞, 普通名詞, サ変可能, *, *, *	左右	サユウ	0	
する	動詞, 非自立可能, *, *, サ行変格, 終止形-一般	為る	する	スル	0
。	補助記号, 句点, *, *, *, *	。	。	0	
EOS					

図 1.1 形態素解析の例

表 3.1 単語分散表現 (word2vec) の例

単語	単語分散表現 (300 次元のベクトル)			
すだち	0.07975651	0.08931299	- 0.06070593	...
人工知能	- 0.14870904	0.02491546	- 0.10277775	...
ai	0.11989793	0.07972154	- 0.26867044	...
データ	- 0.26566359	0.37290227	- 0.16432133	...

2 国語研日本語ウェブコーパス [https://pj.ninjal.ac.jp/corpus\\_center/nwjc/](https://pj.ninjal.ac.jp/corpus_center/nwjc/) (アクセス日: 2020年1月31日)

### 3.2.1 データマイニング

データマイニング (Data Mining) は、大規模なデータベースを高速に解析するための技術や、そこから得られた統計データから重要な指標を計算するための統計技術等からなる。これらは教師なし学習と呼ばれる機械学習の一種として説明されることも多い。近年は、教師あり学習を用いたものも多くみられるが、本研究では、トピック抽出および時系列分析に分類されるテキストマイニングにより分析を行う。

- 判別・クラシファイア (Classifying)
  - 異常検出 (Anomaly Detection)
  - 分類・クラスタリング (Clustering)
  - トピック抽出 (Topic Extraction)
  - 時系列分析 (Time Series Analysis)
- 教師あり学習
- 教師なし学習

### 3.2.2. 時系列分析

時系列分析は、時間の経過に沿って記録された時系列データが対象となる。本研究では、日経新聞と Google Trend を対象に、キーワードの出現頻度の時系列データを対象とした傾向変動 (トレンド) を読み解く。

## 4 ◆ 新聞記事を用いたテキストマイニング

本研究では、日経新聞の 2000 年から 2018 年までの一面<sup>3</sup>の「見出し」と「本文」を対象に「人工知能」をトピックの代表キーワードとする人工知能トピックについて、関連キーワードを抽出し、そのトピックに含まれるキーワードの変遷を時系列データとして分析する。

### 4.1. 日経新聞のデータ

今回入手した記事データは、表 4.1 で示すように 13 の項目に区切られている。日本経済全体における「しごと」の変化を見出すことが目的のため、比較的大きなイベントについて記述されることが多い headline (見出し) と bodysub (本文) を用いる。記事データは、図 4.1 の記事データサンプルのように CSV 形式 (UTF-8) にて提供されており、headline 及び bodysub は、どちらも日本語のテキストデータで提供されている。

### 4.2. キーワードの分析

headline 及び bodysub は日本語のテキストデータで提供されているため、イベント情報を分析するためには、まずキーワードに分割する必要がある。今回は、分かち書きに形態素解析器 Sudachi を使い、さらにその単語分散表現として、国語研日本語ウェブコーパスで学習した word2vec のモデルを利用する。

### 4.3. 人工知能トピックに含まれるキーワード

人工知能トピックに含まれるキーワードは、キーワード間の類似度が高いものとする。本研究では、キーワード間の類似度としてコサイン類似度 (cos 類似度; cosine) を用いる。例えば、「人工知能」と「人間」の類似度を計算するためには、まず「人工知能」の単語分散表現と「人間」の単語分散表現を取得し、ベクトルで表される 2 つの単語分散表現間の角度を類似度とする。このとき単語分散表現間の角度は cosine と可換であるため、キーワード間の類似度は、-1 から 1

3 日本経済新聞社提供の記事データ (日経朝刊の 1 面記事, 2000 年~2018 年, 約 4 万 5 千件)

表 4.1 日経新聞の記事データに含まれる項目

項目名	内容	備考
articleid	記事 ID	"NIRKDB20150715NKM0351" 等
date	日付	"19830316" 等
media	媒体名	"日本経済新聞 朝刊" 等
mediacode	媒体略号	媒体のユニークコード
men	面名	地方経済面の場合に収録。"名古屋朝刊社会面"、"埼玉" 等
pagefrom	ページ	掲載ページ。但し、痴呆経済面のページ情報は、掲載されたページではなく、どの地方の記事であるかを意味する。
picture	絵・写真・表の有無	"有" or "有"
paragraph	段落数	
length	文字数	
headline	見出し	
bodysub	本文	
keyword	キーワード	記事の文中から主題語として切り出したワード (またはその正式名称)
bunrui	分類語	記事内容のテーマコード (#W ~)・業界コード等の会社コード (T ~、N ~、PD ~)、紙面名等の記事分類キーワード (\$ ~)、コラム名 (「~」)

**記事データ例**

```

articleid,date,media,mediacode,men,pagefrom,picture,paragraph,length,headline,bodysub,keyword,bunrui
"NIRKDB20150808NKM0326","20150808","日本経済新聞 朝刊","NKM","","15","","4","406","マブチ純利益4%
減、今期174億円、自動車向け伸びる、予想より改善。","マブチモーターは7日、2015年12月期の連結純利益が前期
比4%減の174億円になる見通しだと発表した。従来予想(11%減の161億円)から減益幅が縮小する。米国で新車生産
台数が増え、主力の自動車用モーターが伸びる。年間配当は105円とし、従来計画から6円引き上げた。今年1月に1株を
2株に分割しており、前期の配当実績は年214円だった。売上高は15%増の1410億円と従来予想を80億円上回る。
日米欧、中国の自動車メーカーにパワーウィンドーやパワーシート、パーキングブレーキ用など主力の中型のモーターが好調だ。
ドアロック用など小型モーターも増える。営業利益は30%増の220億円を見込む。利益率の高い中型モーターの販売増が
寄与する。ドル高・円安が進み輸出採算も改善し、売上高営業利益率は15・6%(従来予想は14・5%)に上向く。今期下
期の為替レートは1ドル=118円(従来想定は108円)に見直した。","マブチモーター,配当,連結決算,純利益,減益,改善,年
間配当,売上高,営業利益,収益見通し,円相場,決算","T6592,PD233,$投資情報,#W10502,#W10501,#W10
408,#W10401,#W30301,#W30303,#B0150,#B0160,#B0170,N0017155"

```

凡例) 書誌項目、見出し、本文、キーワード、分類

図 4.1 日経新聞の記事データサンプル

の範囲の実数で表される cos 類似度で表す。キーワード  $a$  とキーワード  $b$  のベクトル表現を  $\vec{a}$  と  $\vec{b}$  としたとき、キーワード  $a$  とキーワード  $b$  の類似度は次式とする。

$$\cos(\vec{a}, \vec{b}) = \frac{\vec{a} \cdot \vec{b}}{|\vec{a}| |\vec{b}|} \quad (4.1)$$

**4.3.1 関連キーワード**

国語研日本語ウェブコーパス NWJC を Sudachi で分かち書きし、word2vec で学習したモデルで、「人工知能」との cos 類似度順の上位 100 件のキーワードを表 4.2 に列挙する。最上位の「人口知能」は「人工知能」の誤字であることに注意されたい。なお、「人工知能」に「しごと」を奪われるといわれている「人間」は 67 番目、cos 類似度は 0.5164 である。「ai」は「AI」を小文字に正規化したものである。

「人工知能（AI）の進化を受け、データの活用法は企業の競争力を左右する。」



(分かち書き⇒分散表現)

'人工知能'	[ 0.11989793 0.07972154 -0.26867044 ... 0.00160107 0.03451492 0.18664911]
('	[-0.03064764 -0.02055943 -0.02918551 ... 0.07980221 -0.1386897 0.21873805]
'ai'	[ 0.11989793 0.07972154 -0.26867044 ... 0.00160107 0.03451492 0.18664911]
...	
'左右'	[ 0.15070561 0.27088812 -0.01537065 ... 0.03071502 -0.13339861 0.19570848]
'する'	[-0.04262996 0.08501831 0.01364794 ... 0.00316689 0.04561372 0.30977321]
'。'	[-0.02831401 0.10125325 -0.12704873 ... 0.35899675 0.03712214 0.02951937]

図 4.2 分かち書きとキーワードの分散表現

表 4.2 「人工知能」の関連キーワード上位 100 件

順位	キーワード	cos 類似度
-	人工知能	1.0000
1	人口知能	0.7196
2	人工生命	0.6926
3	ロボット工学	0.6657
4	人工無脳	0.6418
5	ロボット	0.6304
6	知能	0.6219
7	コンピューター	0.6060
8	人工無能	0.5995
9	ai	0.5957
10	人間そっくり	0.5917
11	コンピュータ	0.5916
12	人間型ロボット	0.5875
13	ヒューマノイド	0.5861
14	スーパーコンピュータ	0.5846
15	二足歩行ロボット	0.5794
16	機械学習	0.5760
17	自然言語処理	0.5692
18	介護ロボット	0.5668
19	認知科学	0.5648
20	人工臓器	0.5607
21	ロドニー・ブルックス	0.5604
22	情報工学	0.5562
23	量子コンピューター	0.5537
24	バーチャルリアリティ	0.5530
25	インタラクション	0.5509
26	頭脳	0.5504
27	スーパーコンピューター	0.5502
28	自然言語	0.5491
29	科学技術	0.5472



> 0.55

30	人造人間	0.5471
31	小型ロボット	0.5469
32	音声認識	0.5468
33	無脳	0.5449
34	コンピュータビジョン	0.5449
35	チューリング	0.5435
36	人型	0.5396
37	バイオモルフ	0.5393
38	ナノマシン	0.5385
39	コンピュータ言語	0.5375
40	量子コンピュータ	0.5374
41	仮想現実	0.5372
42	コンピュータシステム	0.5345
43	情報処理	0.5344
44	アンドロイド	0.5306
45	知識工学	0.5299
46	工学	0.5296
47	科学者	0.5295
48	ボット	0.5287
49	演算能力	0.5283
50	ロボット兵器	0.5278
51	ロボティクス	0.5269
52	uicn	0.5263
53	思考実験	0.5262
54	未来予測	0.5259
55	ロボット三原則	0.5249
56	ピュータ	0.5235
57	人工知能学会	0.5212
58	ロボットの時代	0.5197
59	人工物	0.5191
60	プログラミング言語	0.5186
61	コンピュータサイエンス	0.5178
62	音声合成	0.5176
63	思考ルーチン	0.5175
64	機械工学	0.5174
65	知的生命体	0.5172
66	脳科学	0.5171
67	人間	0.5164
68	未来社会	0.5151
69	女性型	0.5141
70	電子工学	0.5136
71	人工心臓	0.5132
72	ロボットアーム	0.5131
73	メカトロニクス	0.5130
74	ドロイド	0.5119
75	計算能力	0.5118
76	自動制御	0.5113
77	エキスパートシステム	0.5113



78	軍用ロボット	0.5108
79	自動操縦	0.5103
80	フォン・ノイマン	0.5103
81	人間について	0.5100
82	コンピュータゲーム	0.5090
83	チューリングテスト	0.5089
84	複雑系	0.5087
85	artificialintelligence	0.5083
86	制御理論	0.5078
87	人工言語	0.5073
88	技術的特異点	0.5058
89	コンピュータグラフィックス	0.5057
90	家庭用ロボット	0.5055
91	人工意識	0.5046
92	生命体	0.5039
93	ニューラル	0.5039
94	マービン・ミンスキー	0.5036
95	仮想世界	0.5034
96	データマイニング	0.5031
97	ンピューター	0.5031
98	ニューラルネット	0.5022
99	コンピュータネットワーク	0.5021
100	システム工学	0.5015

#### 4.4 時系列データの分析

人工知能トピックの時系列データの分析では、2000年1月から2018年12月までの日経新聞一面の見出しと本文に含まれるキーワード「人工知能」と、その関連キーワードの出現頻度を時系列にみる。また比較対象として、2004年1月から2018年12月までのGoogle Trendの出力結果をみる。

##### 4.4.1 時系列データの分析

人工知能の関連キーワードを、「人工知能」との関連度が高いキーワード（cos類似度 > 0.55）としたとき、2000年から2018年の日経新聞一面の見出しと本文には、「人工知能」の関連キーワードが10語（表4.2の下線キーワード）含まれ、図4.3のグラフのようになった。

2004年ごろまでは、人工知能トピック全体の頻度は徐々に低下している。2000年の内訳は図4.4のように72%が「コンピューター」であったが、2004年には図4.5のように42%まで低下しており、トピック全体の低迷は「コンピューター」の頻度の低下が要因であるといえる。その後しばらくはほとんど増減がない。ILSVRCでHintonらがDeep Learningを用いて画像認識コンテストで優勝した2012年（Krizhevsky, Sutskever and Hinton 2012）から2013年にかけても、トピック全体の総数は変わらず、その内訳も図4.6からほとんど変化がみられない。ところが、2014年ごろから「ロボット」の出現頻度が伸び始め、それに追随するように「ai」や「人工知能」の頻度が増加する。2015年からは「機械学習」が初登場し、2018年までかけて徐々に登場回数が増加する。さらに、2016年から2018年にかけては、「人工知能」と「ai」の頻度は爆発的に伸び、図4.7のように「ai」「人工知能」「ロボット」の3つのキーワードが人工知能の関連キーワードの94%を占める状態となった。

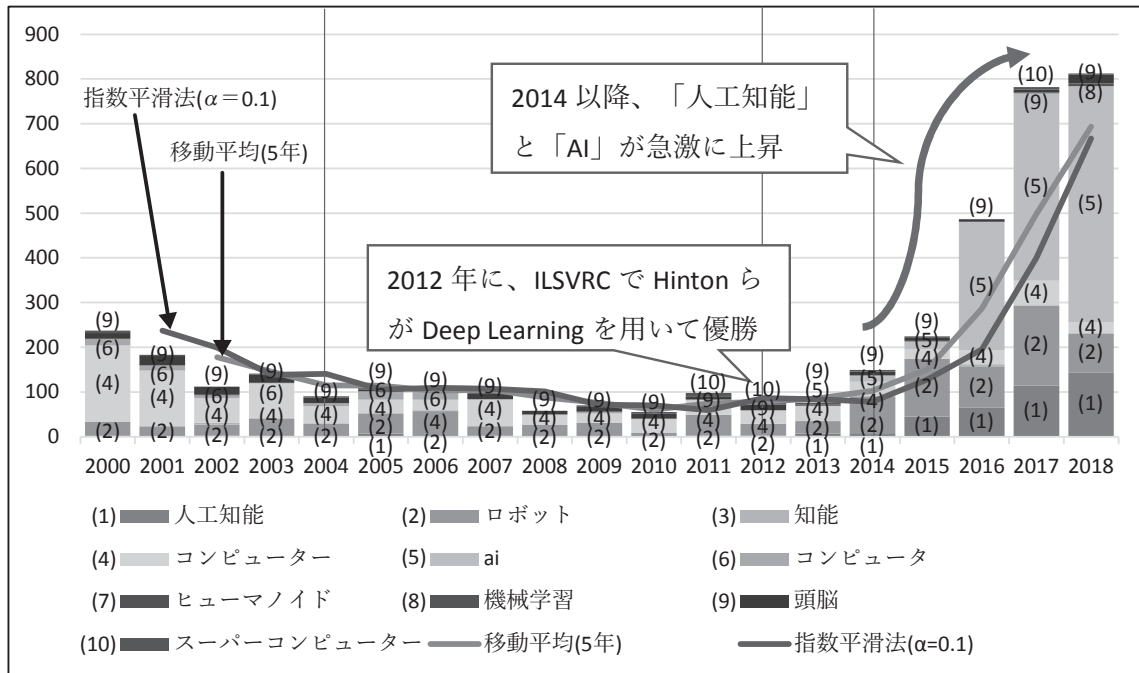


図 4.3 人工知能トピックのトレンド

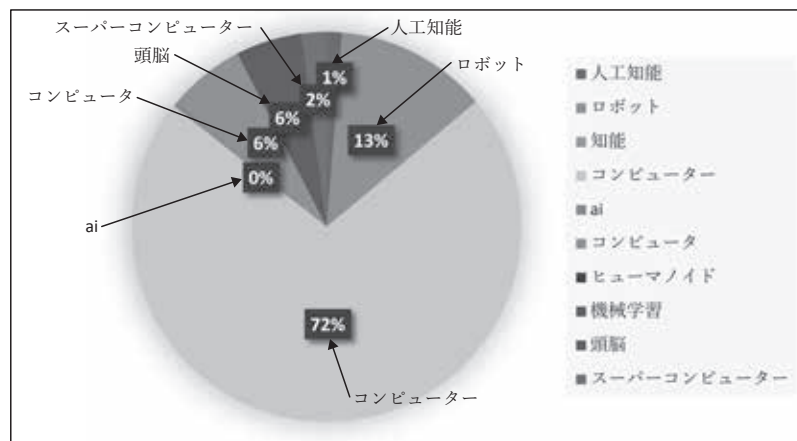


図 4.4 2000 年のキーワードトレンド (日経新聞一面)

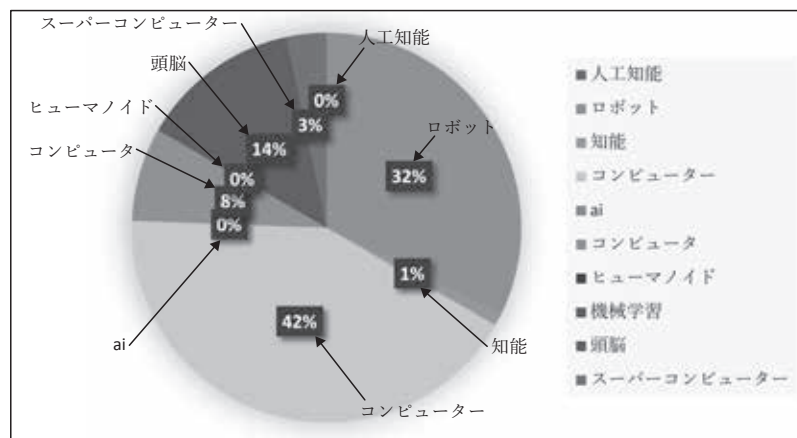


図 4.5 2004 年のキーワードトレンド (日経新聞一面)

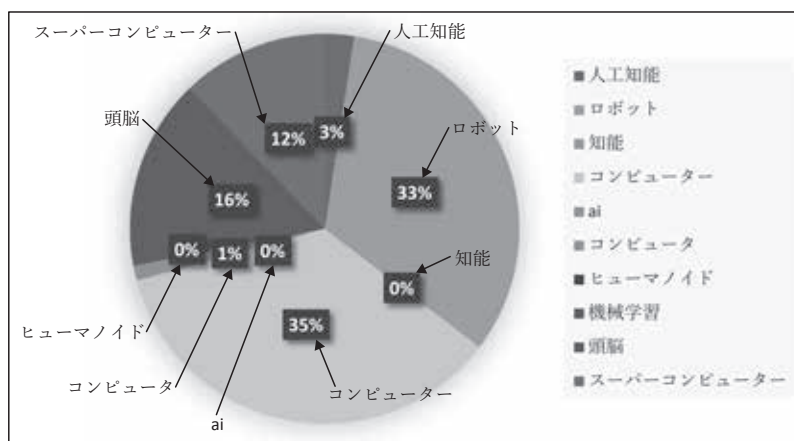


図 4.6 2012年のキーワードトレンド (日経新聞一面)

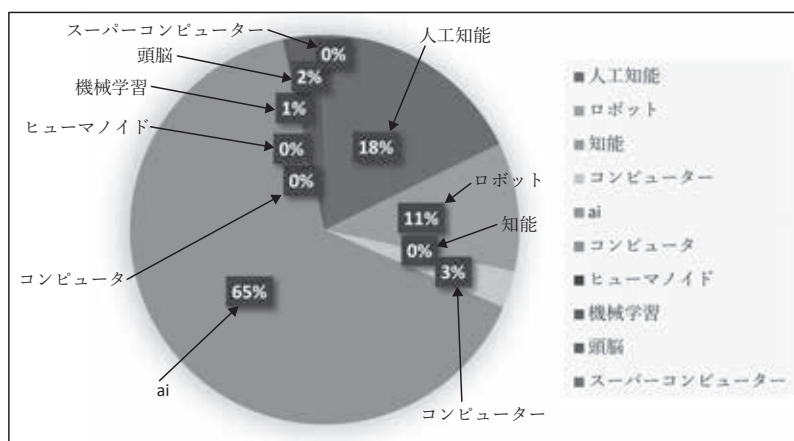


図 4.7 2018年のキーワードトレンド (日経新聞一面)

#### 4.4.2. Google Trend を用いた分析

Google 社が提供する Google Trend<sup>4</sup> を用いることで、2004 年 1 月以降の検索キーワードの頻度を元にしたトレンドをみることができる。図 4.8 は、人工知能トピックの関連キーワードのうち「AI」「ロボット」「知能」「コンピュータ」「機械学習」の 5 語を検索キーワードとして、2004 年 1 月から 2018 年 12 月までのトレンドをグラフで表している。キーワードのスコアは人気度 (Popularity) であり、平均的な人気度のキーワードのスコアを 50 として、そのスコアの範囲は 0 から 100 で表される。

日経新聞一面との違いは、2000 年から 2003 年までのデータがないことである。そのため、2000 年から 2003 年までの間の「人工知能」や「AI」関連のスコアについては不明である。

図 4.8 をみるとわかるように、2004 年の時点では、「ロボット」や「コンピュータ」のスコアが高いが、2012 年までかけて徐々に低下する (図 4.9)。ただし、「ロボット」のスコアについては、2012 年以降も高いスコアを維持している (図 4.10)。これに対し、「コンピュータ」のスコアは 2012 年以降も低下し、これと入れ替わる形で「AI」のスコアが伸びている (図 4.11)。また、

4 Google Trends <http://trends.google.com/trends> (参照日: 2020 年 1 月 31 日)

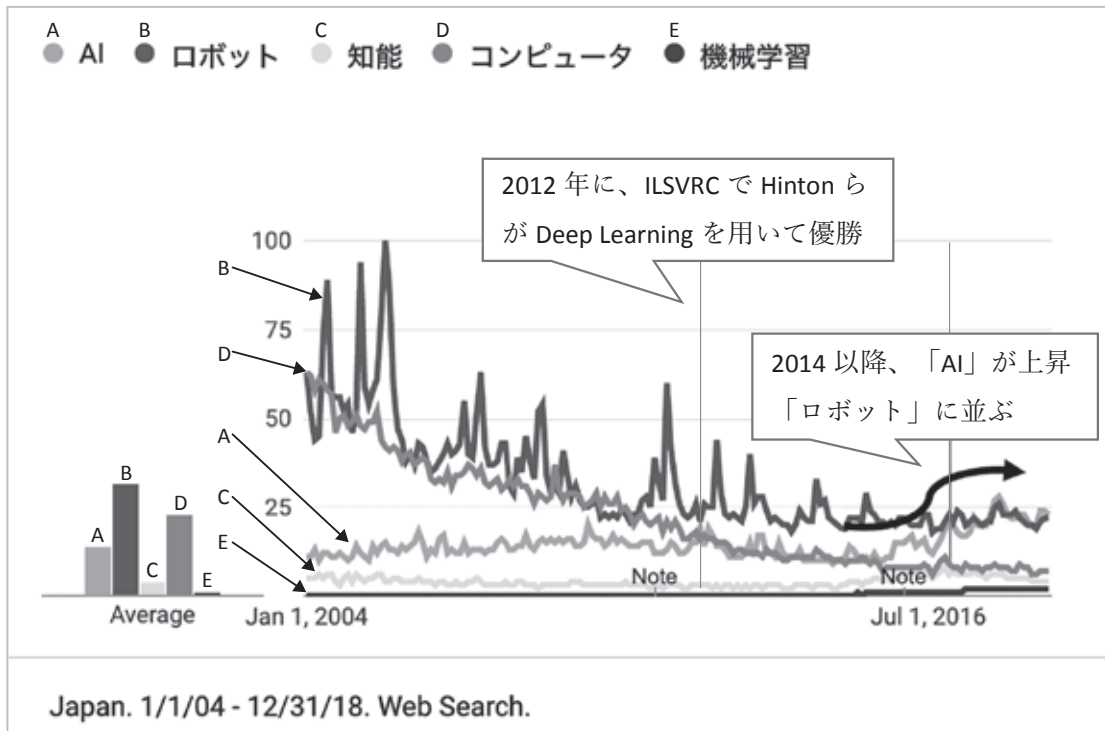


図 4.8 Google Trend を用いた AI のトレンド

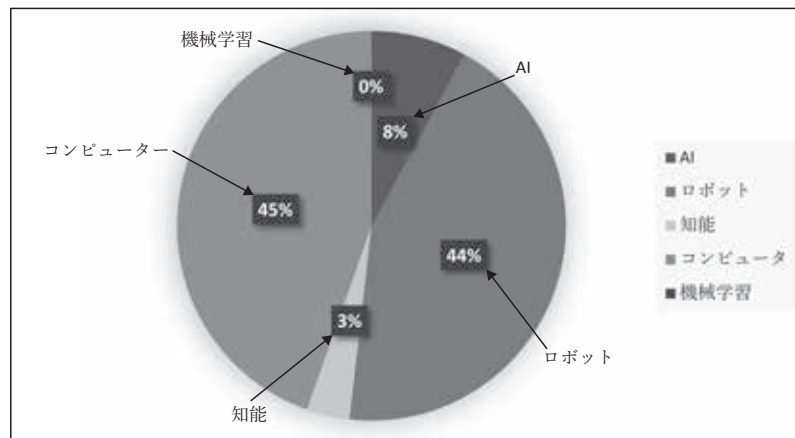


図 4.9 2004年1月のキーワードトレンド (Google Trend)

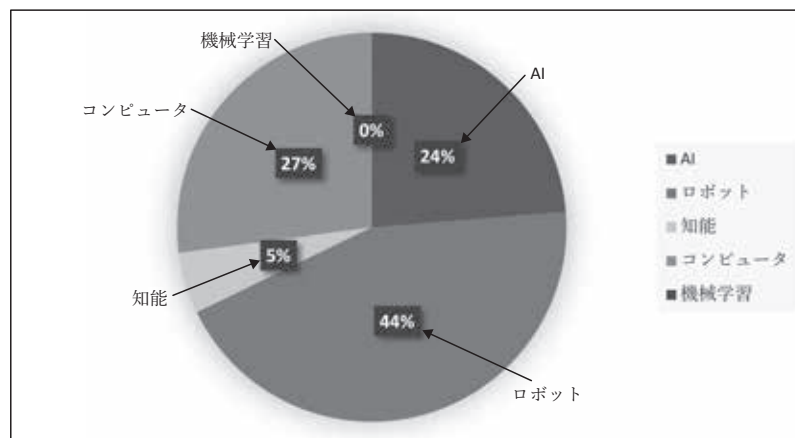


図 4.10 2012年1月のキーワードトレンド (Google Trend)

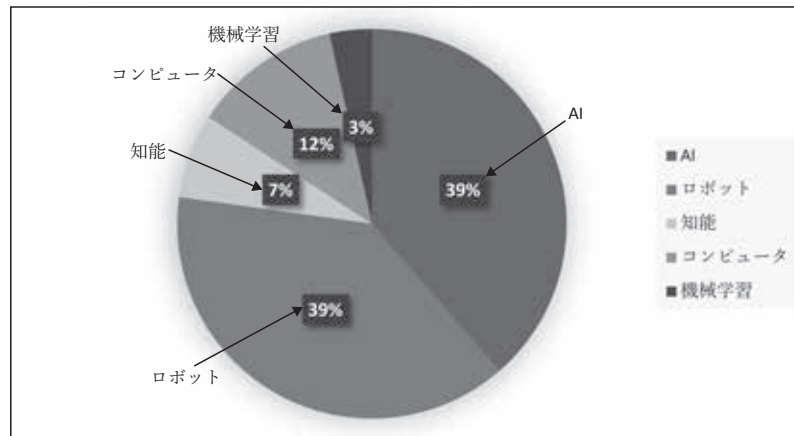


図 4.11 2018年12月のキーワードトレンド (Google Trend)

2015年ごろから「機械学習」のスコアが少しずつ上昇しており、日経新聞一面と同じ傾向を示している。

## 5◆考察

本研究では、日経新聞一面で「人工知能」や「AI」が2000年から2018年にかけて、どのように取り上げられたか、2004年から2018年にかけて、Googleの検索クエリとしてどのように人気があったかについて確認し、その傾向を明らかにした。その結果、新聞記事とGoogle Trendともに2012年ごろがターニングポイントとなっており、2014年ごろからはさらに「人工知能」や「AI」の注目度が伸びていることがわかった。

特に、2016年以降に日経新聞一面で取り上げられた頻度の伸びには特筆すべきものがある。これはGoogle Trendと比較しても異常値を示しているといわざるをえない。また、表2.1のように「しごと」の増加・減少について言及された時期とも一致する。この頃に、人工知能が新たな労働力として注目され始めたため、日本経済が敏感に反応したといえる。

## 6◆まとめと今後について

本稿では、日経新聞およびGoogle Trendを用いて“人工知能”が2012年ごろから徐々に注目され始めたことを明らかにした。また、2016年以降の日経新聞では「人工知能」と「AI」の露出が爆発的に伸びており、Google Trendと比較しても、企業を中心とした日本経済全体での人工知能への関心の大きさがわかる。この状況を、ガートナー社が提唱したハイブ・サイクル(ハイブ曲線; hype cycle) (Fenn 1995) に当てはめると、まさに流行期を迎えているといえる。つまり今後、幻滅期を迎え、第3次AIブームは終焉を迎える可能性があるといえる。それでもなお、人工知能の技術の発達を前提として、教育や産業の改革を推し進めようとするのであれば、我々はその改革の歩みを止めてはならない。我々にはその覚悟が必要である。

今後の課題としては、このAIブームが日本全体へどのように波及しているかについて、地方紙や日経新聞の地方面を利用することで明らかにすること。さらに、現在のAIブームが幻滅期を迎え、かつての第2次AIブームの足跡を辿るか否か、または幻滅期を迎えた後も人工知能は産業構造の中心技術となっていけるか否かについて、追跡調査をすることがあげられる。

## 参考文献

- ・伝康晴・小木曾智信・小椋秀樹・山田篤・峯松信明・内元清貴・小磯花絵 (2007) 「コーパス日本語学のための言語資源：形態素解析用電子化辞書の開発とその応用」日本語科学 第22巻, 101-123 ページ。
- ・真鍋陽俊・岡照晃・海川祥毅・高岡一馬・内田佳孝・浅原正幸 (2018) 「複数粒度の分割結果に基づく日本語単語分散表現」『言語処理学会 第25回年次大会 発表論文集』言語処理学会, 1407-1410 ページ。
- ・経済産業省 (2016) 「第4次産業革命 人材育成推進会議 資料5」アクセス日：2020年1月31日。  
[http://www.kantei.go.jp/jp/singi/keizaisaisei/miraitoshikaigi/jinzaikusei\\_dai1/siryou5.pdf](http://www.kantei.go.jp/jp/singi/keizaisaisei/miraitoshikaigi/jinzaikusei_dai1/siryou5.pdf)
- ・Blei, David M., Andrew Y. Ng, and Michael I. Jordan (2003) “Latent Dirichlet Allocation,” *Journal of Machine Learning Research*, JMLR.org, 3, 993-1022.
- ・Deerwester, Scott, Susan T. Dumais, George W. Furnas, Thomas K. Landauer, and Richard Harshman (1990) “Indexing by latent semantic analysis,” *Journal of the American Society for Information Science*, John Wiley & Sons, Inc., 41 (6), 391-407.
- ・Fenn, Jackie (1995), *When to Leap on the Hype Cycle*. Gartner Group, Accessed : 31 January 2020. <https://www.wordspy.com/words/hypecycle.asp>.
- ・Krizhevsky, Alex, Ilya Sutskever, and Geoffrey E Hinton (2012) “ImageNet Classification with Deep Convolutional Neural Networks,” Edited by F. Pereira, C. J. C. Burges, L. Bottou and K. Q. Weinberger. *Advances in Neural Information Processing Systems 25*, Curran Associates, Inc., 1097-1105.
- ・Mikolov, Tomas, Kai Chen, Gregory S. Corrado, and Jeffrey A. Dean (2013) “Computing numeric representations of words in a high-dimensional space,” United States Patent, US9037464B1.
- ・Takaoka, Kazuma, Sorami Hisamoto, Noriko Kawahara, Miho Sakamoto, and Yoshitaka Uchida (2018) “Sudachi : a Japanese Tokenizer for Business,” Edited by Nicoletta Calzolari, Khalid Choukri, Christopher Cieri, Thierry Declerck, Sara Goggi, Koiti Hasida, Hitoshi Isahara, et al., *Proceedings of the Eleventh International Conference on Language Resources and Evaluation (LREC 2018)*, Paris, France. European Language Resources Association (ELRA).