BANK OF JAPAN

Indexing and Visualization of Climate Narratives Using BERT and Causal Extraction

4th IFC and Bank of Italy Workshop
on “Data Science in Central Banking”

2025/2/19

Bank of Japan Noriyasu Kaneda
The Hokkaido university Hiroki Sakaji

Background

- *"Climate change is a global challenge and could have a broad impact on our society and economic activity into the future."*

  \- BOJ's Strategy on Climate Change

- Economists in central banks are focusing on how climate change will affect financial markets and macro-economy.

- Some researches used traditional text analysis to capture climate-related risk from newspaper and creating climate index using the NLP methods.

  Engle et al. [2020], Pastor et al. [2022], Faccini et al. [2023], Hiraki et al. [2025]

Overview and Contribution

- We proposed a method to analyze cause-effect relationship in text data (we call it "economic narrative"), using two NLP techniques, (1) BERT and (2) causal extraction.

  (1)BERT (Bidirectional Encoder Representations from Transformers) is a Deep Learning based language model by Google, and it can learn the sentence structure and context in a text.

  (2)We also use another NLP, "Causal Extraction" (Sakaji et al. [2008]). This method is an algorithm based on linguistic knowledge.

- This research focuses on climate narratives. We analyze the newspapers on climate addressing by using the NLP methods above, and index/visualize climate narratives.

Data

【 Nikkei newspaper (morning edition, weekdays only) 】

- Period: January 2000 - November 2021
- 17,000 climate change-related articles* in Japanese

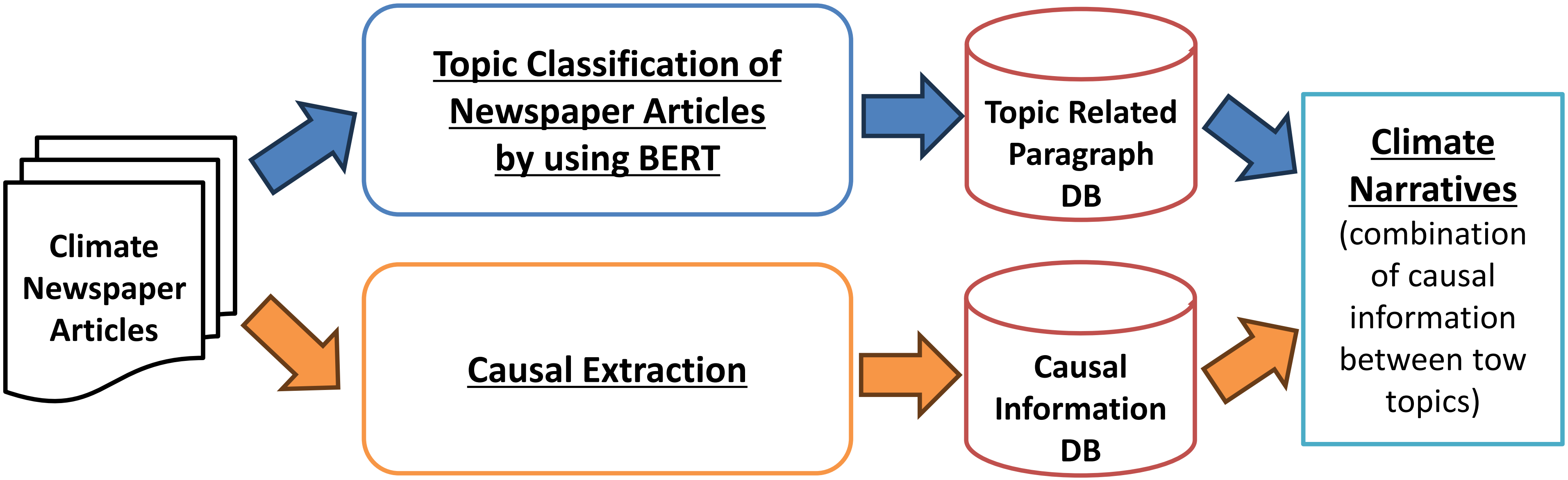
*containing at least one of the words "climate change," "global warming," or "greenhouse effect.

■ 40 topics (classification tags) are the target of this analysis.

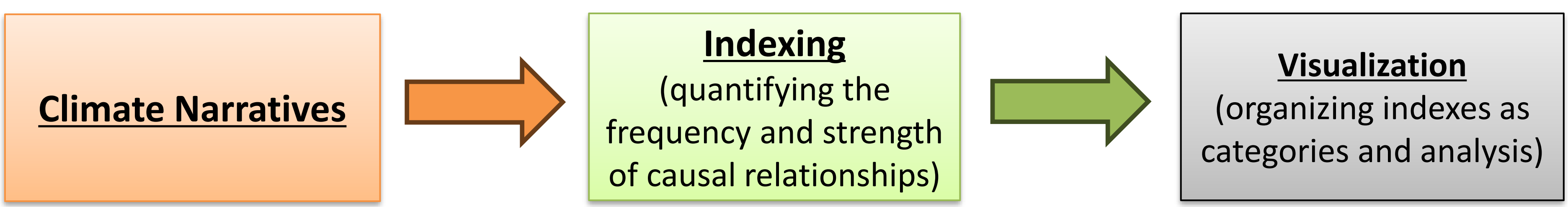
【Corporate】		【Politics】	【Economics】	【Society】
1 New business	11 Governance	21 Party	28 Monetary policy	35 Weapon
2 M&A	12 Labour	22 Regulation	29 Inflation	36 Disaster
3 Business strategy	13 Name change	23 Fiscal policy	30 Business cycle	37 Trial
4 Price strategy	14 Wage	24 Energy policy	31Finacial market	38 Energy probrem
5 Production strategy	15 Finance	25 Seculity	32 Foreign Exchange	39 Environmental probrem
6 Cost reduction	16 Performance	26 Summit	33 Bond	40 Consumer trend
7 Supply chain	17 Sales	27 Social seculity	34 Renewable energy	
8 Patent	18 Price			
9 R&D	19 Market share			
10 Investment	20 Hot selling			

Methodology for Extracting Climate Narratives

Methodology for Extracting Climate Narratives

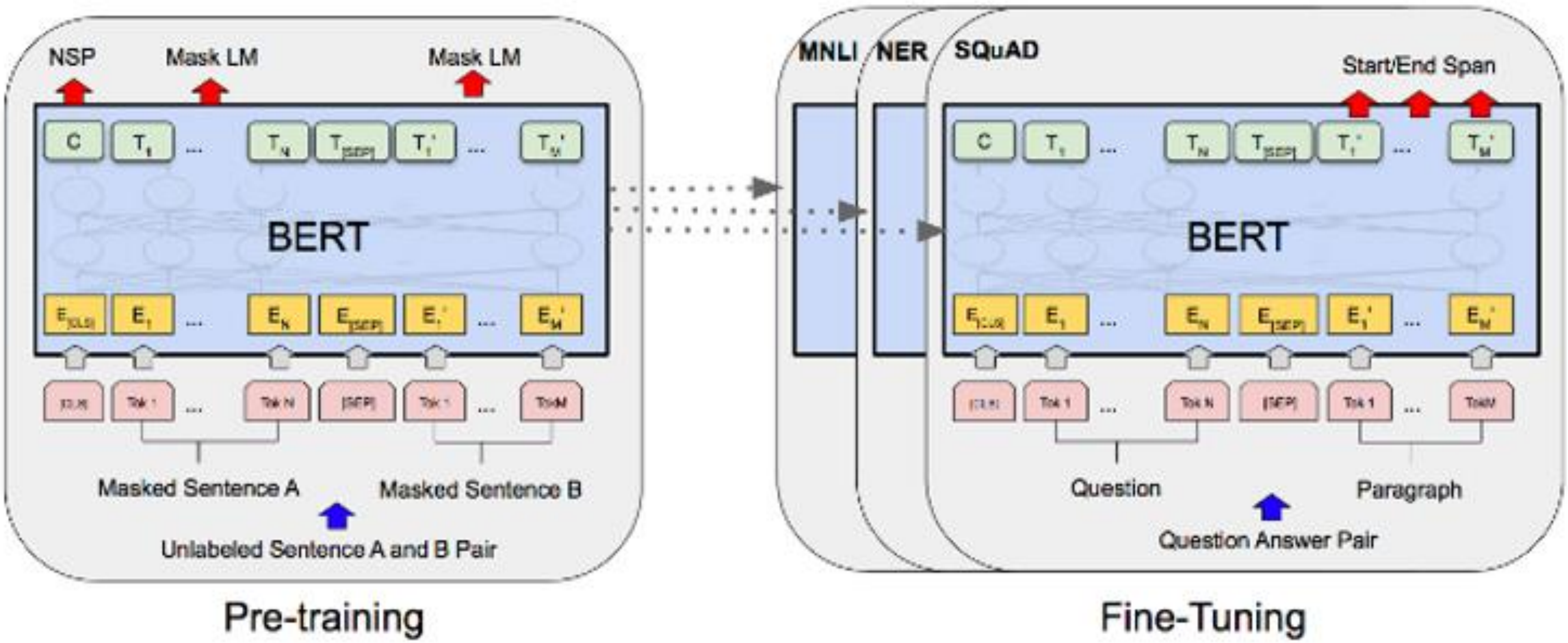


Methodology for Indexing and Visualization of Climate Narratives (for empirical research)



BERT (Bidirectional Encoder Representations from Transformers)

Devlin et al. [2019]



Trained and fine-tuned by Google for treating basic linguistic problems (public source)

More Fine-tuned for classification of climate news by topics (specific purpose)

Results of the "OLYMPIC" Classification Model

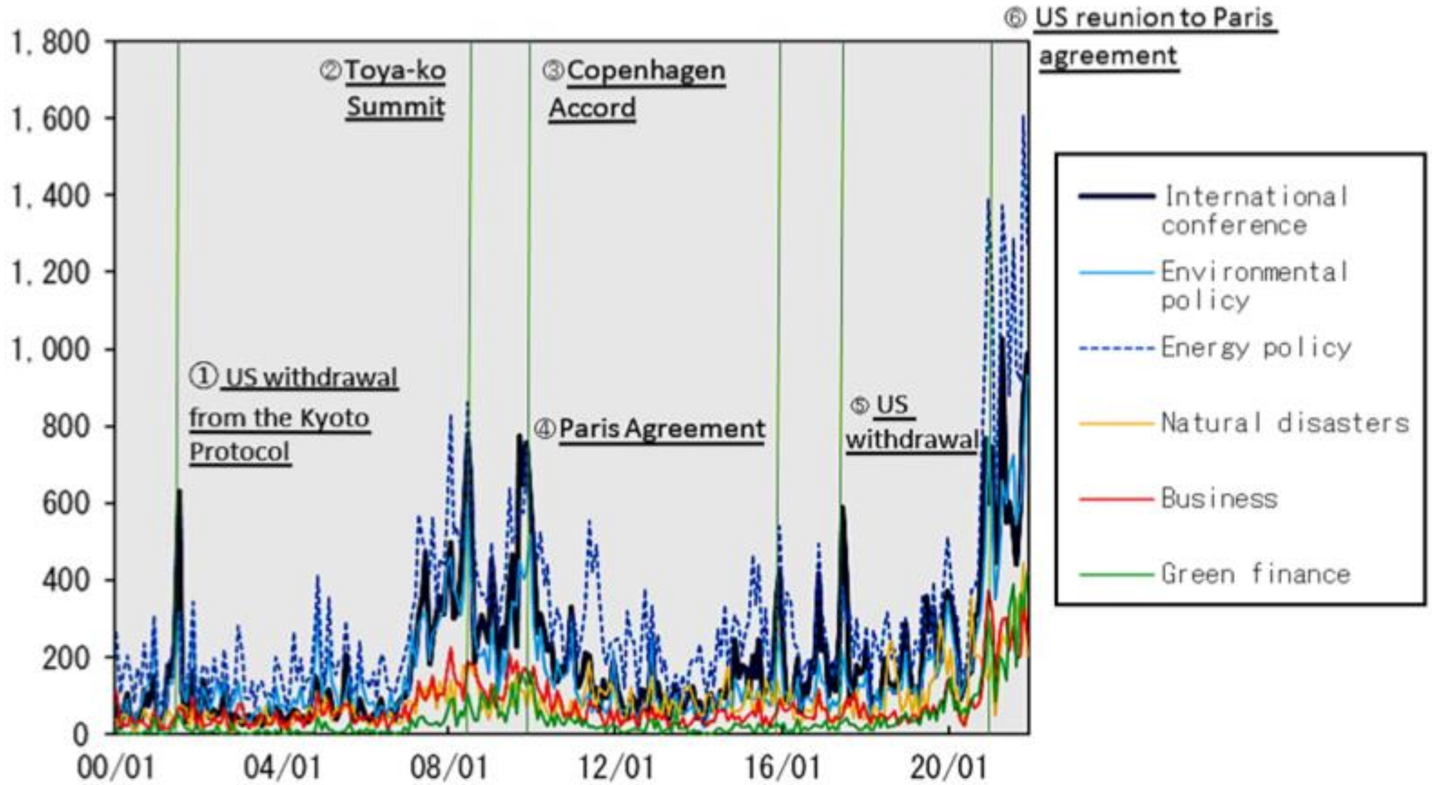
	Accuracy rate	Precision rate	Recall rate	F1 score
Linear Regression	0.958	0.958	0.958	0.958
Random Forest	0.950	0.951	0.950	0.950
Support Vector Machine	0.954	0.954	0.954	0.954
BERT	0.960	0.960	0.960	0.960

- In the classification for Olympic article as an example, BERT is slightly better than other models.

Basic analysis (Topic based index using only BERT)

- **Topic based index** increased through 2008 to 2010 (in Toya-ko Summit, Copenhagen Accord) and in 2015(the Paris Agreement).

- In recent years (since 2018), many topics show a surge, including micro developments such as corporate strategy.



⇒ High correlation among topics. It is not clear how climate narratives are linked across topics.

Causal extraction: The second NLP for finding causality

- We focus on linguistic causal-effect relationship and consider the causal relationships in newspaper as economic narratives by using causal extraction.

Step.1 Causality Determination [Sakaji et al. 2017]

- Determine whether or not the input sentence contains a causal relationship (a causal sentence).

Step.2 Causality Extraction [Sakaji et al. 2008]

- Extract "cause and effect expressions" from causal sentences that contain causal relations picked up in Step1.

- This method analyzes syntactic patterns based on dependency parsing in linguistic knowledge.

Example of Climate Narrative (from Regulation to Corporate topic)

Causal sentence 1 on Environmental Regulation in Jan 2016

With Paris agreement agreed to in 2015, climate change policies are being discussed around the word.
As a result, the Japanese government plans to strengthen the environment to promote decarbonization policies.

High cosine similarity (similar information)

Causal sentence 2 on Corporate Strategy in Feb 2017

In response to the government's move to tighten regulations, corporates are begging to incorporate climate change addressing into their corporate strategies.

We observe a climate narrative from Regulation to Corporate in Feb 2017.
If there are many narratives in Feb 2017, this narrative is a hot topic at this point.

Generalization

<u>Causal Event (Topic A)</u>

Causal expression ⇒ ***Effect expression***

Prior causal information (past)

High cosine similarity

Combine a pair of "Causal event's *effect expression*" and "*effect expression's causal expression": A result relates Topic B is caused by some event relates Topic A.*

<u>Effect Event (Topic B)</u>

Causal expression ⇒ Effect expression

Subsequent causal information (present)

Indexing Economic Narratives

- Links between Cause $(\vec{i}_{t-d})$ <past information from t time to d days ago> and Effect $(\vec{j}_t)$ <information as of time t> are calculated based on the cosine similarity (how close the textual information is in content) of the causal chain and sum it by month.

$$Index\_monthly\_m = \sum_{j=0}^{M} \sum_{i=0}^{L(j)} \frac{1}{1+ae^{bd}} cos(\vec{i}_{t-d} \cdot \vec{j}_t) \quad (1)$$

$$cos(\vec{i}_{t-d} \cdot \vec{j}_t) = \frac{\vec{i}_{t-d} \cdot \vec{j}_t}{|\vec{i}_{t-d}||\vec{j}_t|} \quad (2)$$

Here,

M : set of causal chains included in month m.

$L(j)$: set of cause event $\vec{i}_{t-d}$ connected to result event $\vec{j}_t$.

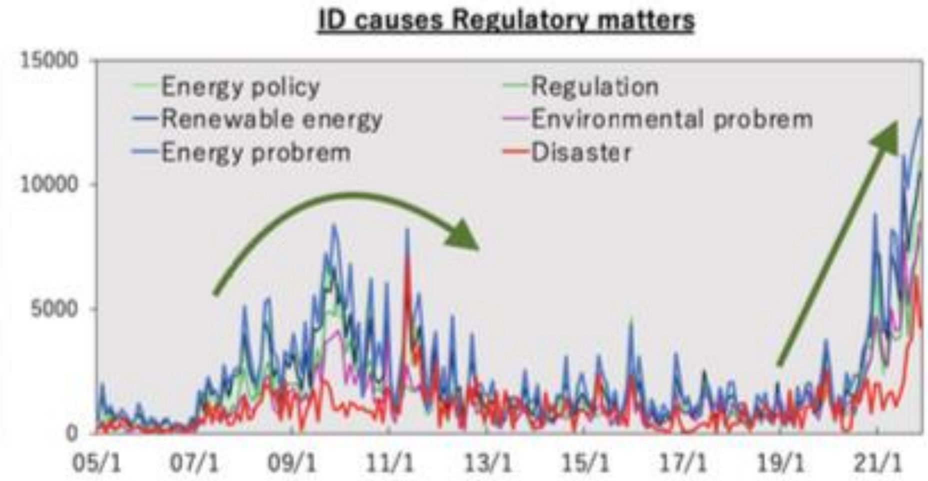
$t-d$: observation point of cause event leading to result event ($d > 0$).

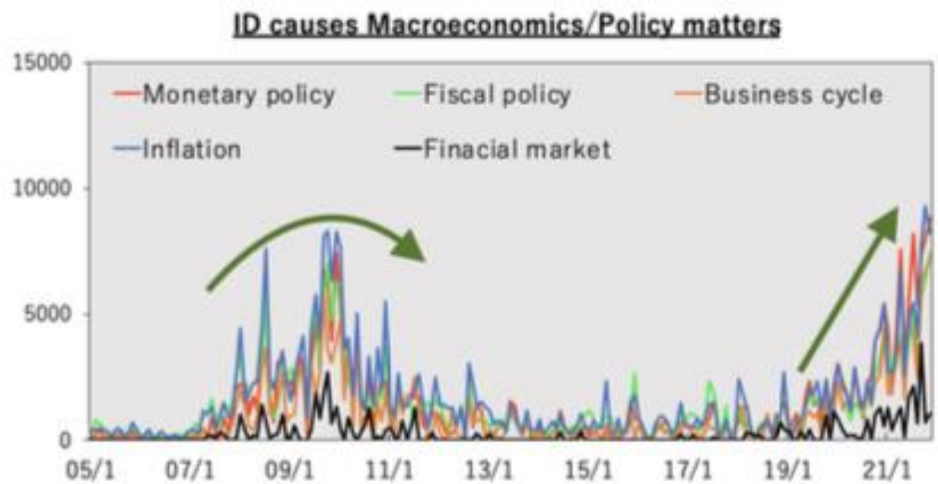
t : observation point of result event included in month m.

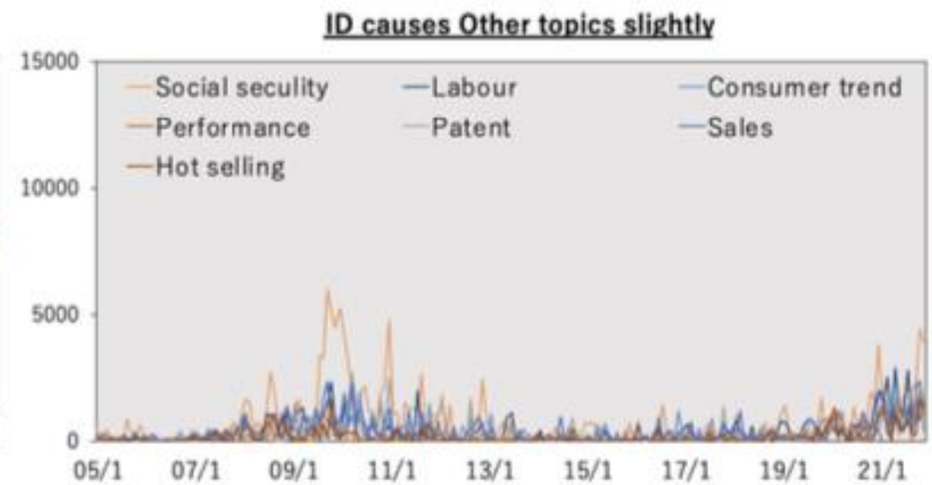
d : time difference (days) between cause event and result event.

*Old causal chains are depreciated according to elapsed time to eliminate the bias that the latest news has more links to the past (weights are halved after 5 years based on the logistic function) in (1). **Count all links with $cos(\vec{i}_{t-d} \cdot \vec{j}_t) \geq 0.7$** (strong causality) in (2).

Climate indexes (International discussion ⇒ each topics)







Mapping all narratives from 2001 to 2017

- Strong links to many topics from international conferences on climate change, environmental policy, and energy policy.

Effect Topic

Cause Topic

		1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37	38	39	40
1	new business		116	203	257	282	272	192	133	237	192	100	147	55	143	98	91	201	409	268	118	190	768	886	464	556	885	265	377	666	279	70	97	84	181	364	430	73	233	198	94
2	M&A	12		32	35	32	29	30	19	45	19	17	18	9	20	18	14	26	58	27	14	32	77	120	98	82	115	39	58	101	45	14	20	18	51	57	52	14	91	59	15
3	business strategy	229	112		247	267	214	185	133	198	234	93	145	55	141	98	94	200	429	275	124	172	859	886	530	521	868	248	364	654	269	62	89	74	171	352	441	65	168	173	91
4	price strategy	135	51	108		132	132	156	61	172	173	43	69	25	65	46	47	86	96	124	51	85	389	397	280	230	392	116	182	251	127	35	48	41	176	159	183	34	168	204	40
5	production strategy	223	102	184	219		233	98	118	272	110	84	125	49	123	93	83	159	305	222	103	162	784	775	439	489	738	224	363	530	239	66	94	73	172	321	382	54	272	185	80
6	cost reduction	435	172	367	375	440		418	208	510	355	153	220	80	222	157	134	284	618	437	168	320	[illegible]	[illegible]	704	836	[illegible]	434	588	963	415	113	151	133	336	558	606	116	407	371	132
7	supply chain	790	272	736	584	740	939		317	831	901	228	332	114	318	241	199	441	852	704	257	483	[illegible]	[illegible]	761	[illegible]	[illegible]	624	920	[illegible]	637	168	240	194	314	829	976	167	623	274	217
8	patent	47	32	62	62	78	85	73		47	92	34	39	15	38	35	23	44	103	57	27	67	106	214	163	145	196	76	110	171	73	21	33	27	94	100	95	28	152	105	28
9	R&D	316	130	262	300	385	446	323	153		437	109	160	58	155	121	100	231	498	344	145	205	900	981	454	602	958	286	404	726	295	68	101	84	155	415	489	79	250	102	119
10	investment	503	200	485	448	486	502	452	233	594		167	249	91	238	174	157	325	698	496	193	345	[illegible]	[illegible]	623	853	[illegible]	462	648	[illegible]	458	124	171	146	266	625	727	115	383	336	159
11	governance	32	11	31	24	30	28	54	16	38	41		12	5	14	13	8	16	41	24	9	31	58	75	109	73	96	33	39	71	28	11	15	14	83	46	36	12	82	76	12
12	labour	60	27	57	57	46	38	94	36	85	42	27		14	29	33	24	37	105	53	25	59	239	188	231	148	210	68	102	147	68	25	34	29	152	100	89	22	172	209	25
13	name change	1	2	4	7	3	1	4	4	7	1	3	2		4	3	3	6	11	4	3	6	21	20	17	14	22	7	11	17	9	4	3	3	4	10	9	3	10	14	3
14	wage	77	24	62	49	72	72	172	33	81	96	26	26	11		26	21	35	86	54	17	58	237	151	254	129	195	45	85	85	53	22	28	31	220	81	78	24	154	244	22
15	finace	23	13	39	31	48	44	81	19	46	67	15	21	7	22		12	23	45	37	12	34	25	114	148	80	132	44	47	102	42	14	18	16	100	48	41	14	116	49	12
16	performance	32	12	27	24	20	17	53	17	46	20	12	13	8	16	14		18	44	25	13	25	142	109	105	69	112	37	52	82	33	12	17	14	75	43	41	9	75	111	12
17	sales	112	36	89	72	70	108	181	43	142	71	31	47	19	46	33	35		119	70	33	62	379	302	194	180	293	89	137	185	86	27	39	33	122	115	135	20	81	108	29
18	price	344	105	360	191	293	507	339	117	412	507	85	141	42	127	90	85	178		280	100	192	655	731	474	515	714	244	353	408	250	79	105	86	378	321	373	63	492	246	88
19	market share	164	82	160	165	162	230	193	94	206	183	74	104	37	98	77	70	119	246		71	135	567	503	545	391	553	187	258	383	182	59	85	65	237	257	277	50	306	245	67
20	hot selling	55	19	26	37	60	73	91	24	55	86	16	28	13	23	20	19	29	62	26		30	203	148	240	87	149	42	67	101	49	12	18	15	120	55	71	12	73	60	18
21	party	125	41	101	74	111	209	313	64	127	242	52	96	19	58	46	29	52	138	107	26		364	175	409	219	290	124	194	221	125	47	61	57	498	182	138	58	452	445	50
22	regulation	[illegible]	416	[illegible]	346	[illegible]	[illegible]	[illegible]	532	[illegible]	[illegible]	407	524	147	512	374	287	616	[illegible]	[illegible]	338	972		[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	322	431	362	[illegible]	[illegible]	[illegible]	341	[illegible]	[illegible]	367
23	fiscal policy	731	228	652	414	669	[illegible]	[illegible]	301	788	[illegible]	247	293	85	287	205	159	304	708	606	146	599	[illegible]		[illegible]	[illegible]	912	620	718	843	559	227	294	240	[illegible]	840	699	214	[illegible]	[illegible]	224
24	energy policy	[illegible]	402	[illegible]	399	[illegible]	[illegible]	[illegible]	503	[illegible]	[illegible]	360	513	157	485	344	295	675	[illegible]	[illegible]	393	846	[illegible]	[illegible]		[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	983	267	384	315	532	[illegible]	[illegible]	283	970	280	371
25	seculity	401	131	348	203	369	613	773	181	421	706	136	152	48	131	109	79	157	398	346	80	376	640	570	806		152	371	536	717	346	109	160	122	923	372	389	135	[illegible]	527	134
26	summit	727	232	654	392	669	[illegible]	[illegible]	296	786	[illegible]	235	279	83	244	195	150	297	692	608	149	636	931	964	[illegible]	807		649	827	[illegible]	586	200	271	219	[illegible]	785	714	208	[illegible]	823	221
27	social security	181	52	154	110	159	288	438	78	190	334	68	77	22	81	64	42	75	183	151	36	141	395	121	665	334	465		176	156	138	58	73	71	660	219	159	65	540	671	55
28	monetary policy	278	95	250	175	278	437	671	125	295	494	108	133	35	124	84	76	131	273	240	61	246	560	84	968	907	485	237		167	173	102	126	103	930	357	268	86	886	851	92
29	inflation	540	151	498	304	471	791	[illegible]	196	577	900	158	208	59	203	139	120	228	441	415	117	369	[illegible]	392	[illegible]	828	925	371	384		321	150	188	160	[illegible]	575	487	124	[illegible]	962	146
30	business cycle	198	64	188	123	171	291	416	77	212	294	65	87	25	77	58	54	87	175	143	46	146	557	155	561	347	375	152	80	61		65	82	68	511	228	193	47	487	511	61
31	finacial market	30	12	26	24	33	50	76	14	30	59	18	18	6	16	12	10	16	28	26	7	33	69	32	120	59	64	35	6	23	29		11	11	122	39	27	10	121	149	11
32	forex	51	20	45	36	51	90	125	24	53	94	22	28	8	23	19	17	26	51	45	12	52	109	27	216	92	67	55	14	32	35	17		21	218	68	48	17	206	263	20
33	bond	38	16	31	27	40	62	104	18	39	80	18	22	5	20	15	12	19	43	38	9	44	59	12	159	87	82	39	3	27	25	14	17		164	54	35	14	154	178	14
34	renewable energy	[illegible]	440	[illegible]	975	[illegible]	[illegible]	[illegible]	548	[illegible]	[illegible]	396	561	175	533	371	326	732	[illegible]	[illegible]	426	901	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	294	421	345		[illegible]	[illegible]	299	[illegible]	295	398
35	weapon	263	84	250	143	243	406	414	111	285	415	81	100	32	80	67	54	109	246	226	56	220	548	440	411	156	222	220	336	466	216	66	97	74	470		277	78	567	363	88
36	disaster	392	85	390	210	362	532	743	109	419	602	70	108	35	100	66	67	162	380	288	112	181	[illegible]	725	477	481	747	201	293	498	217	48	77	61	525	351		62	486	126	117
37	trial	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]		[illegible]	[illegible]	[illegible]
38	energy probrem	[illegible]	401	[illegible]	915	[illegible]	[illegible]	[illegible]	502	[illegible]	[illegible]	356	515	162	499	343	300	682	[illegible]	[illegible]	401	803	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	965	261	372	311	534	[illegible]	[illegible]	270		297	361
39	environmental probrem	[illegible]	533	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	677	[illegible]	[illegible]	493	685	203	651	456	389	868	[illegible]	[illegible]	505	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	381	544	435	[illegible]	[illegible]	[illegible]	390	[illegible]		502
40	consumer trend	52	15	48	26	47	76	115	19	61	96	16	18	6	18	14	10	21	49	36	15	37	141	66	154	72	91	38	48	63	39	14	19	14	104	52	43	14	107	50	

International Conference

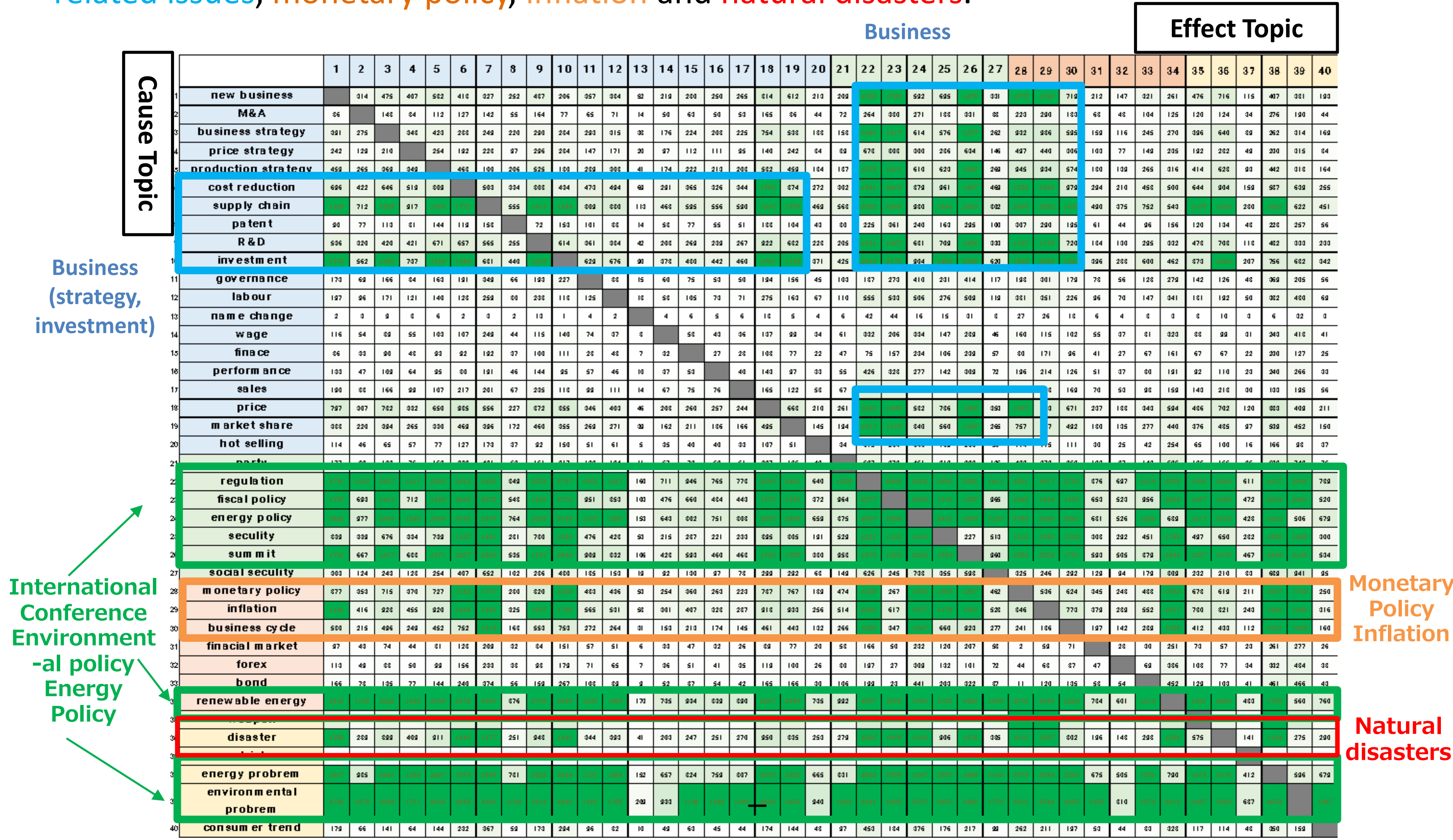
Environment-al policy

Energy Policy

Note: Dark green highlight means a high level of economic narrative index between each topic.

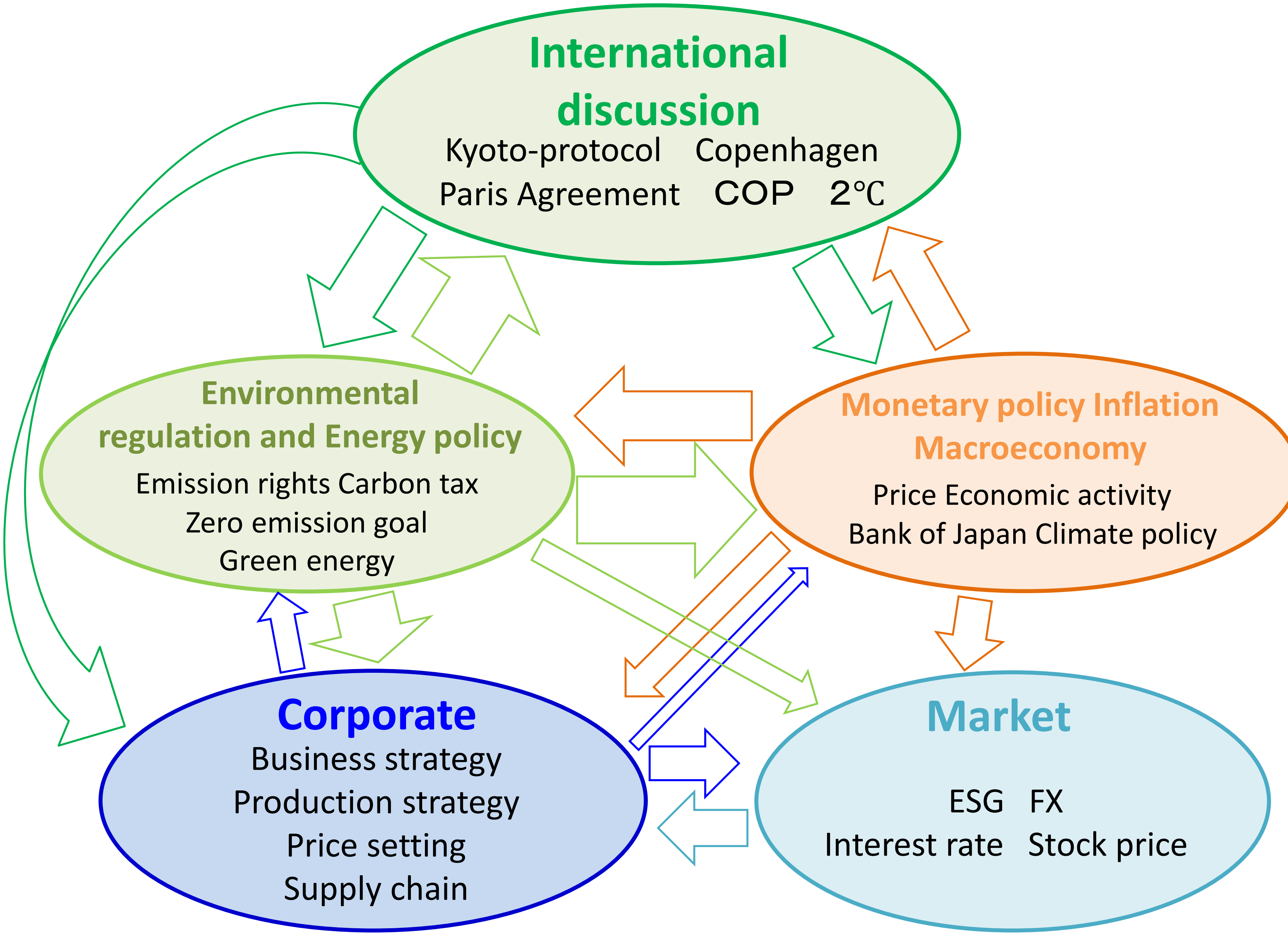
Mapping all narratives from 2018 to 2021

- Strong connections on a wide range of topics from international conferences on climate change, environmental policy and energy policy. Connections have emerged in business-related issues, monetary policy, inflation and natural disasters.



Note: Dark green highlight means a high level of economic narrative index between each topic.

Visualization of Climate Narratives



Note: The size of an arrow means a level of economic narrative index between each topic categories.

Main Result

- The climate narratives suggest that only governments played a major role in the discussion of international regulations for carbon neutrality in the post-Kyoto period, or before the Paris Agreement.

- In recent years, climate change has become a hot topic not only in international and domestic regulations, but also in corporate strategy, green finance and big issues, and as a major issue for central banks since 2018.

- These developments could be related to the growing awareness of transition risk and physical risk in Japan.

Discussions

➢ Is climate narrative from causal sentence statistical causality?

✓ Newspaper could contain noisy or incorrect information due to the misunderstanding of the article writers.

✓ Empirical analysis of market data using climate narrative indexes is the next step (like Engle et al. [2020], Faccini et al. [2023]).

➢ A possible next topic: how central bank narratives about climate risk have affected the behavior of other economic entities?

"*An economic narrative is a contagious story that can change how people make economic decisions*" (Shiller [2020]).

Reference

- J. Devlin, M.-W. Chang, K. Lee, and K. 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), 2019, pp. 4171–4186.
- R. F. Engle, S. Giglio, B. Kelly, H. Lee, and J. Stroebel, "Hedging climate change news," The Review of Financial Studies, vol. 33, no. 3, pp. 1184–1216, 2020.
- R. Faccini, R. Matin, and G. Skiadopoulos, "Dissecting climate risks: Are they reflected in stock prices?" Journal of Banking & Finance, vol. 155, p. 106948, 2023.
- K. Hiraki, N. Kaneda, T. Kimata, and T. Mtsue, "Climate change news indices: Are they reflected in japanese stock prices?" , IMES Discussion Paper Series, 2025, forthcoming.
- R. J. Shiller, *Narrative economics: How stories go viral and drive major economic events.* Princeton University Press, 2020.
- L. Pastor, R. F. Stambaugh, and L. A. Taylor, "Dissecting green returns," Journal of Financial Economics, vol. 146, no. 2, pp. 403–424, 2022.
- H. Sakaji, S. Sekine, and S. Masuyama, "Extracting causal knowledge using clue phrases and syntactic patterns," in Practical Aspects of Knowledge Management. Springer Berlin Heidelberg, 2008, pp. 111–122.
- H. Sakaji, R. Murono, H. Sakai, J. Bennett, and K. Izumi, "Discovery of rare causal knowledge from financial statement summaries," in 2017 IEEE Symposium Series on Computational Intelligence (SSCI), 2017, pp. 602–608.