

“Visualization” of SDGs-related Patent Technologies by Natural Language Processing Technology

Yoshiaki MAEHARA¹, Atsushi KUKU¹ & Yoshiyuki OSABE¹¹ Japan Patent Information Organization, Tokyo, JAPAN, <https://transtool.japio.or.jp/work/en/>

Correspondence: Yoshiyuki OSABE, Japan Patent Information Organization, Tokyo, JAPAN

Received: August 3, 2021

Accepted: August 18, 2021

Online Published: August 20, 2021

doi:10.11114/bms.v7i3.5315

URL: <https://doi.org/10.11114/bms.v7i3.5315>

Abstract

SDGs is an abbreviation for “Sustainable Development Goals”, which were adopted at the United Nations Sustainable Development Summit in September 2015. They were set as goals to be achieved by the 193 United Nations Member States during the 15-year period from 2016 to 2030. Technological innovation is indispensable for the realization of the SDGs, but at present, it is unclear where (countries and companies) and to what extent SDG-related technologies are available. For this reason, we used BERT, a natural language processing technology, and Japanese patent publications we own and worked on the "visualization" of the SDGs technologies in Japan. The results show that out of the 17 SDGs goals, patents can contribute to 2, 3, 6, 7, 9, 11 and 13.

Keywords: SDGs, ESG, BERT, NLP, Patent

1. Introduction

The whole world has been tormented by climate change.

After a Canadian village experienced a high temperature of 49.6 degree, a wildfire burned the entire village to the ground. In Germany and Belgium, floods swept through many cities like a tsunami. In Zhengzhou, China, almost a year's amount of rain fell in just three days, causing extensive flooding. In Japan as well, record-breaking torrential rains in 2018 caused river flooding, inundation, and landslides mainly in western Japan, resulting in an enormous disaster with more than 200 fatalities. Now, it has been predicted that, according to climate change, the temperature and the number of extreme weather events would rise and increase gently, but in fact this is not the case, and we are now entering an era of "non-linear" movements on climate change.

Under these circumstances, a lot of attention is being paid to the movement, like SDGs and Decarbonization, to solve problems on a global scale.

SDGs is an abbreviation for “Sustainable Development Goals”, which were adopted at the United Nations Sustainable Development Summit in September 2015. They were set as goals to be achieved by the 193 United Nations Member States during the 15-year period from 2016 to 2030 (United Nations, 2021). Specifically, SDGs include 17 major goals and 169 specific targets to be met to achieve those goals.



Figure 1. Sustainable Development Goals

Patent database is a strategic resource for technical information and utilized in many fields. For example, research and development of new product (Cascini, 2007, Zanni-Merk, 2009, Russo, 2011, Trappey, 2012), emerging technology forecast (Daim, 2006, Trappey, 2011), patent quality analysis (Ernst & Omland, 2011, Jibu & Osabe, 2014, 2015), problem solving (Cascini, & Russo, 2007, Verhaegen, 2011, Zanni-Merk, 2011), etc. Recently, it has also been used as IP landscape for evidence-based management strategy (Jibu & Osabe, 2018).

In order to achieve full competence in patent searching, patent classifications are useful tools (White, 2010). Each IP Office has issued its own patent classification, for example, International Patent Classification (IPC) by WIPO, FI and F-term by JPO, ECLA, ICO, and Tagging Codes by EPO, DeKla by DPMA, and USPC by USPTO (Wolter, 2012). In recent years, EPO and USPTO have jointly developed a new patent classification (CPC: Cooperative Patent Classification). However, no IP office has developed a classification related to SDGs.

In an article, SDG relevant innovation is quantified with a keyword search method in patent texts (Johannes, 2021). However, the analysis is limited because it is done by a keyword search method, comprehensive analysis has not been conducted.

Although there have been several previous studies on SDGs-related text classification using machine learning (Zhang, 2020, Medina, 2019), there has been no previous study on classifying patent documents from the perspective of SDGs. In other words, technological innovation is indispensable for the realization of the SDGs, but at present, it is unclear where (countries and companies) and to what extent SDG-related technologies are available.

In this study, we used Japanese patent publications and BERT (Devlin, 2018), a natural language processing technology, and worked on the "visualization" of the SDGs technologies existing in Japan. BERT, which stands for "Bidirectional Encoder Representations from Transformers", is a famous transformer for producing the best results with the smallest corpus. It is also the model that produces the highest accuracy in the patent classification task (Lee, J. S. & Hsiang, J., 2019, Zaheer, M., 2020).

2. Method

2.1 Analytical Model

In this study, we used BERT, a pre-trained language model, as the machine learning model. BERT is a bidirectional transformer trained by two tasks: Masked Language Model and Next Sentence Prediction. When applied to specific tasks such as sentiment analysis and text classification, very high accuracy can be achieved with a small corpus.

Firstly, we created a model by adding a 17-class classification layer to the bottom layer of the Japanese BERT (Communication Science Laboratory, Tohoku University, 2019) developed and distributed by Tohoku University. The task is a single-label task, and the output of the final layer is normalized by the Softmax function.

2.2 Training Data

Since BERT is used in this study, the size of the corpus as training data does not need to be very large. For that reason, 17 goals and the underlying 169 target sentences of the SDGs were adopted as training data.

2.3 Analytical Data

We analyzed the Japanese patent publications owned by our organization. Since the input layer of the Japanese BERT model distributed by Tohoku University has 512 fixed input token sequences, we made BERT read the part of the description of the "problem to be solved by the invention" where the features of the patent are likely to appear.

2.4 Targeted Training and Analysis

Cross entropy was used as the loss function and the model after 100-epoch training was analyzed. Among the estimated probabilities of each label (goal) normalized by Softmax, the label with the highest probability was adopted as the label of the patent document. The accuracy validation of the model after training was not confirmed by cross-validation, but the label accuracy rate of the training data itself was 100%.

3. Results

As a result of assigning labels to all patent publications of patent applications filed after 2010, the distribution for assigning SDGs labels is shown in Fig. 2.

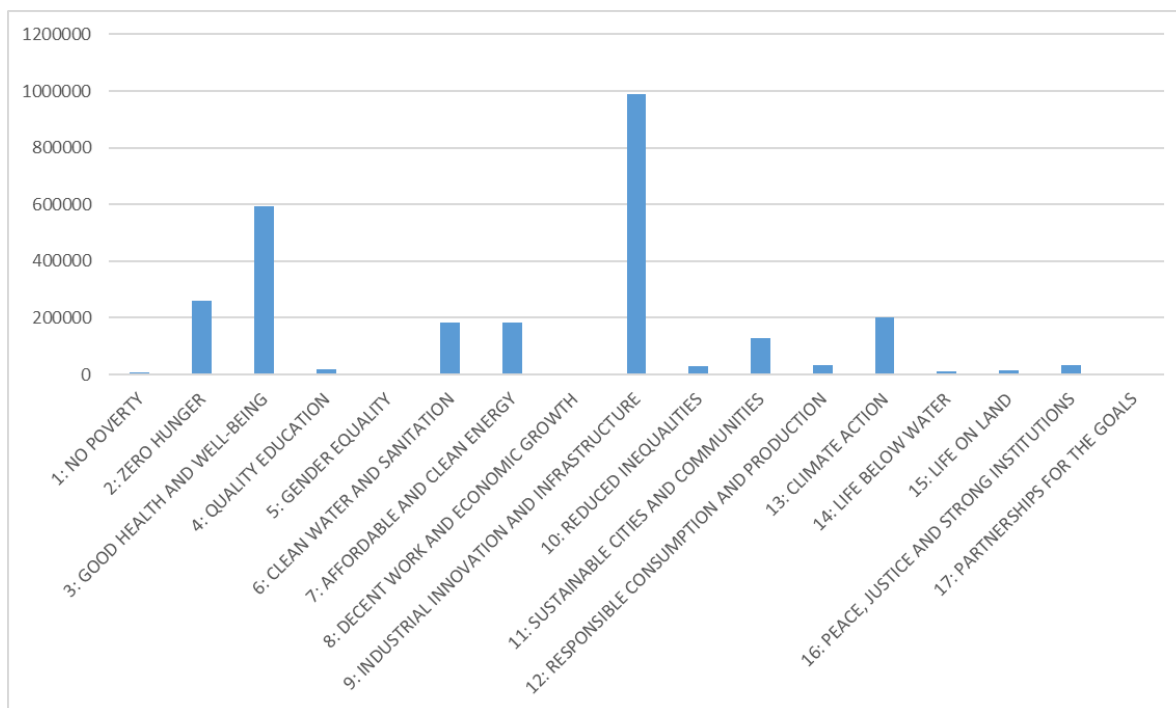


Figure 2. Distribution of SDGs labels (vertical axis) assigned to Japanese patent publications filed after 2010

The results show that out of the 17 SDGs goals, patents can contribute to Goal 2 (ZERO HUNGER), Goal 3 (GOOD HEALTH AND WELL-BEING), Goal 6 (CLEAN WATER AND SANITATION), Goal 7 (AFFORDABLE AND CLEAN ENERGY), Goal 9 (INDUSTRY, INNOVATION AND INFRASTRUCTURE), Goal 11 (SUSTAINABLE CITIES AND COMMUNITIES), and Goal 13 (CLIMATE ACTION). This is broadly in line with the SDGs goals mentioned by WIPO (World Intellectual Property Organization), to which patents can contribute (WIPO, 2021).

Table 1 shows samples of the patents that were extracted as corresponding to each goal.

Table 1. Examples of patent documents corresponding to each SDG

Estimated-SDGs	A part of descriptions of patent documents
1	To provide a system to support customers' credit decisions. ... This system uses the customer information ...
2	It is expected that the productivity of agriculture will be further improved by using robots or IT. ...
3	The invention predicts new functional biomarkers of vascular inflammation and all-cause mortality or cardiac mortality ...
4	To provide a new form of educational service based on knowledge units. ... One embodiment of the invention ...
5	Multiple occupants can perform household work efficiently at the same time and other daily supplies such as clothing ...
6	Increase in usage fees due to aging infrastructure facilities in the future and depletion of water resources caused by global warming ...
7	Hybrid energy supply circuits, use of hybrid energy supply systems and hybrid ...
8	It is said that with the development of AI, many of the jobs of employers will be automated. And merit system ...

9	While controlling the deterioration of the quality of the linked services provided by linking multiple services ...
10	To be able to obtain money equivalent quickly and efficiently to virtual currency. ... Virtual currency trading ...
11	The construction assumptions and the construction unit price magnification corresponding to the assumed disaster of the planned construction site ...
12	To provide a waste sorting apparatus that enables smooth processing of waste containing carbon fiber ...
13	To provide an acclimatization evaluation device that can evaluate the period required for acclimatization quantitatively and in advance, and can be effectively used in advance as a preventive measure against heat stroke...
14	To provide a drifting ocean observation buoy with a small and simple configuration that can control its course. ...
15	To efficiently manage information on the capture of vermin. ... The management system is user's ...
16	To provide highly entertaining fighting games. ... It is a fighting game. The game field ...
17	To improve the overall result of the group compared to methods that give each group a common theme. ...

In order to verify the estimate accuracy of the model, we manually checked the correctness of the estimates in descending order up to 100 patents that were normalized by Softmax for each goal.

For Goals 2, 3, 6, 7, 8, 9, and 11, the results with high estimates were without problems. However, for those that were inappropriate from the perspective of science and technology, such as Goal 1 (NO POVERTY), most of them were noise, even if they are highly estimated. For Goal 4 (QUALITY EDUCATION), patent documents on machine learning were included as noise, influenced by the word "learning". Also, for Goal 13, many patent documents related to air conditioners were included, probably because they were influenced by the word "temperature".

In addition, as a semi-macro analysis, for the Japanese patent publications in 2020, we compared companies that are representative of Japanese telecommunications carriers, KDDI CORPORATION, Nippon Telegraph and Telephone Corporation, NTT DOCOMO, INC., SoftBank Corp., Yahoo Japan Corporation, Z Holdings Corporation, and Rakuten, Inc.(Fig. 3)

While most companies show a high percentage of only Goal 9, Rakuten, Inc. also shows a high percentage of Goal 8. Even within the same industry, companies have different strategies for each of the SDGs goals.

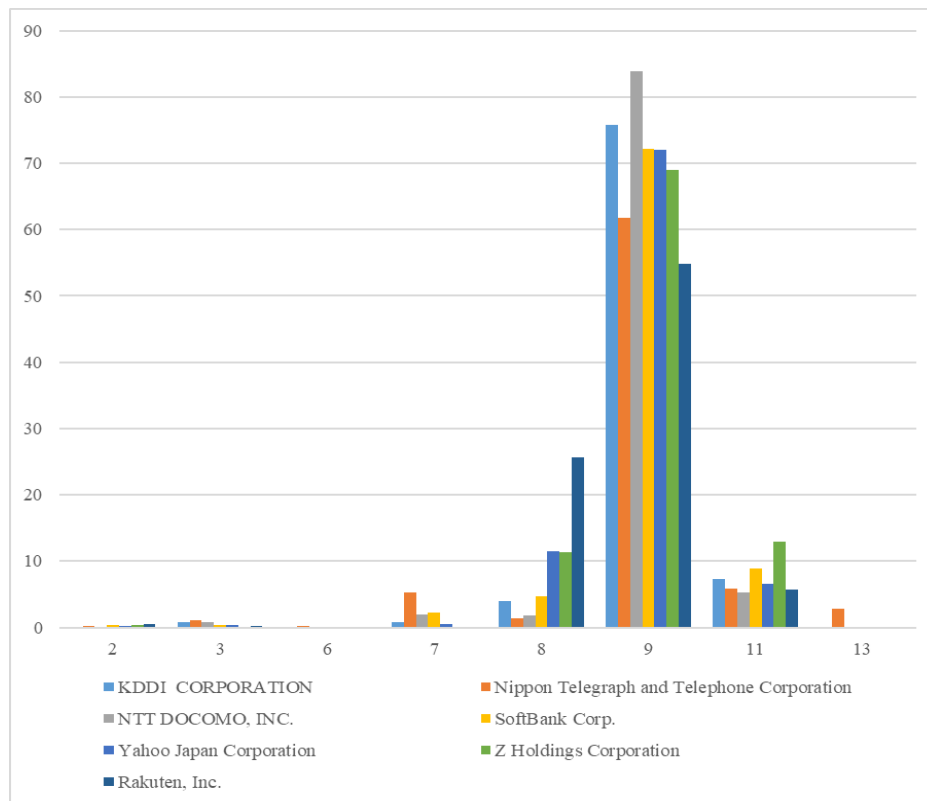


Figure 3. Distribution of applications by Japanese telecommunications carriers (Ratio of the number of Goals 2, 3, 6, 7, 8, 9, 11 and 13 to the total number in the Japanese patent publications in 2020 of each company)

4. Discussion

As shown above, we have succeeded in “visualization” of SDGs-related patent technologies by BERT fine-tuning. In recent years, financial products such as investment trust targeting listed companies working on the SDGs have been actively developed (MSCI Inc., 2020). The statistical indicators generated by this technology will help in the development of those financial products.

In addition, if this technology could be used to search for individual patents related to each SDG, it would help developers and inventors to find collaboration partners. This is because the goal of the SDGs is common means that the problem to be solved by the patent technology is common.

For these applications, it remains an important issue to improve the accuracy of the classification model developed in this study. From now on, we would like to improve the classification accuracy by either increasing the corpus or using BERT dedicated to the patent domain as a model.

References

- Cascini, G. et al. (2007) Computer-aided patent analysis: finding invention peculiarities, in: N. León-Rovira (Eds), *Trends in Computer Aided Innovation*, Springer Boston, 250, 167–178, https://doi.org/10.1007/978-0-387-75456-7_17
- Cascini, G. & Russo, D. (2007) Computer-aided analysis of patents and search for TRIZ contradictions, *International Journal of Product Development*, 4 (1/2), 52–67, Retrieved August 3, 2021 from <http://www.inderscience.com/offer.php?id=11533>
- Communication Science Laboratory, Tohoku University (2019) Pretrained Japanese BERT models released, Retrieved August 3, 2021 from <https://www.nlp.ecei.tohoku.ac.jp/news-release/3284/>
- Daim, T. U. et al. (2006) Forecasting emerging technologies: use of bibliometrics and patent analysis, *Technological Forecasting and Social Change*, 73 (8), 981–1012, <https://doi.org/10.1016/j.techfore.2006.04.004>
- Devlin, J. et al. (2018) BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding. arXiv:1810.04805

- Ernst, H. & Omland, N. (2011) The Patent Asset Index - a new approach to benchmark patent portfolios, *World Patent Information*, 33 (1), 34–41, <https://doi.org/10.1016/j.wpi.2010.08.008>
- Jibu, M & Osabe, Y. (2014) Development of new indicators for the launch of AMED (8): knowledge flows of the pharmaceutical innovation system, *Journal of Information Processing and Management*, 57 (8), 562–572, <https://doi.org/10.1241/johokanri.57.562>
- Jibu, M & Osabe, Y. (2015) Knowledge flows and delays in the pharmaceutical innovation system, Proceedings of ISSI, Istanbul: 15th International Society of Scientometrics and Informetrics Conference, Retrieved August 3, 2021 from https://www.issi-society.org/proceedings/issi_2015/0877.pdf
- Jibu, M & Osabe, Y. (2018) *Scientometrics*, ISBN: 978-1-78923-307-0, Print ISBN: 978-1-78923-306-3, InTech, Retrieved August 3, 2021 from <https://www.intechopen.com/books/scientometrics>
- Johannes W.H. et al. (2021) The innovative contribution of multinational enterprises to the Sustainable Development Goals, *Journal of Cleaner Production*, 285, 125319-125332, <https://doi.org/10.1016/j.jclepro.2020.125319>
- Lee, J. S. & Hsiang, J. (2019) PatentBERT: Patent Classification with Fine-Tuning a pre-trained BERT Model. arXiv:1906.02124
- Medina, S. R. (2019) Multi-Label Text Classification with Transfer Learning for Policy Documents, UPPSALA UNIVERSITET, Retrieved August 3, 2021 from https://github.com/vondersam/sdgs_text_classifier
- MSCI Inc., (2020) MSCI ACWI Sustainable Impact Index, Retrieved August 3, 2021 from <https://www.msci.com/msci-acwi-sustainable-impact-index>
- Russo, D. et al. (2011) Methodological enhancements for concept exploration in product design, *Int. J. Prod. Dev.*, 15 (1/2/3), 28, <https://doi.org/10.1504/IJPD.2011.043660>
- Trappey, C. V. et al. (2011) Using patent data for technology forecasting: China RFID patent analysis, *Advanced Engineering Informatics*, 25 (1), 53–64, <https://doi.org/10.1016/j.aei.2010.05.007>
- Trappey, A. J. C. et al. (2012) A patent quality analysis for innovative technology and product development, *Advanced Engineering Informatics*, 26 (1), 26–34, <https://doi.org/10.1016/j.aei.2011.06.005>
- United Nations, THE 17 GOALS, Retrieved August 3, 2021 from <https://sdgs.un.org/goals>
- Verhaegen, P. A. et al. (2011) Identifying candidates for design-byanalogy, *Computers in Industry*, 62 (4), 446–459, <https://doi.org/10.1016/j.compind.2010.12.007>
- White, M. (2010) Patent Searching: Back to the Future How to Use Patent Classification Search Tools to Create Better Searches, *First Annual Conference of the Canadian Engineering Education Association*, Kingston, Ontario, Retrieved August 3, 2021 from <https://ojs.library.queensu.ca/index.php/PCEEA/article/view/3155/3093>
- WIPO, The Impact of Innovation: WIPO and the Sustainable Development Goals, Retrieved August 3, 2021 from <https://www.wipo.int/sdgs/en/story.html>
- Wolter, B. (2012) It takes all kinds to make a world – some thoughts on the use of classification in patent searching, *World Patent Information*, 34, 8–18, <https://doi.org/10.1016/j.wpi.2011.08.001>
- Zaheer, M. et al. (2020) Big Bird: Transformers for Longer Sequences. arXiv:2007.14062
- Zanni-Merk, C. et al. (2009) An ontological basis for computer aided innovation, *Computers in Industry*, 60 (8), 563-574, <https://doi.org/10.1016/j.compind.2009.05.012>
- Zanni-Merk, C. et al. (2011) Use of formal ontologies as a foundation for inventive design studies, *Computers in Industry*, 62 (3), 323–336, <https://doi.org/10.1016/j.compind.2010.09.007>
- Zhang, X. (2020) Creation of a Japanese SDGs dataset and a baseline model of classification, The Japanese Society for Artificial Intelligence, https://doi.org/10.11517/pjsai.JSAI2020.0_1D3GS1305

Copyrights

Copyright for this article is retained by the author(s), with first publication rights granted to the journal.

This is an open-access article distributed under the terms and conditions of the [Creative Commons Attribution license](https://creativecommons.org/licenses/by/4.0/) which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.