Information Engineering Express
International Institute of Applied Informatics
2024, Vol. 10, No. 2, IEE802

Generating a Technical Trend Map by Analyzing the Structure of U.S. Patents Using Patent Families

Jun Nakamitsu*, Satoshi Fukuda*, Hidetsugu Nanba*

Abstract

Researchers and developers search for patents in fields related to their own research to obtain information on issues and effective technologies in those fields for use in their research. However, it is impossible to read through the full text of many patents, so a method that enables patent information to be grasped briefly is needed. In this study, we analyze the structure of U.S. patents with the aim of extracting important information. Using Japanese patents with structural tags such as "field", "problem", "solution", and "effect", and corresponding U.S. patents (patent families), we automatically created a dataset of 81,405 U.S. patents with structural tags. Furthermore, using this dataset, we conduct an experiment to assign structural tags to each sentence in the U.S. patents automatically. For the embedding layer, we use a language representation model BERT pretrained on patent documents and construct a multi-label classifier that classifies a given sentence into one of four categories: "field", "problem", "solution", or "effect". We are able to classify sentences with precision of 0.6994, recall of 0.8291, and F-measure of 0.7426. We have analyzed the structure of U.S. patents using our method and generated a technological trend map, which confirms the effectiveness of the proposed method.

Keywords: patent, document structure analysis, machine translation, machine learning, technical trend map

1 Introduction

When researchers and company engineers consider new research or development, utilizing patent information is important for grasping the latest technical trends. On the other hand, it is difficult to read through all the patents published around the world. Under such circumstances, a method that enables an efficient overview of technical trends is needed. To overview technical trends, it is effective to classify patents according to the viewpoints of technologies and problems, etc. However, to do so, it is necessary to extract the description part of technologies and problems from each patent. Therefore, this study aims to analyze the structure of U.S. patents.

Unlike U.S. patents, Japanese patents have explicit items such as "Field of Technology" (hereinafter referred to as "field"), "Problem to Be Solved by the Invention" (hereinafter referred to as "problem"), "Solution for Solving the Problem" (hereinafter referred to as "solution"), and

* Chuo University, Tokyo, Japan

5.2 Results

The evaluation results are shown in Table 3. As can be seen from the table, the results of clustering by solution sentences matched the analysis axes almost perfectly. The results of projecting the clustering results using GMM onto a 2-dimensional plane with t-SNE [15] are shown in Figure 1. As can be seen from Figure 1, clustering using solution sentences is the most appropriate way to separate the analysis axes.

Figure 2 shows the abstract of US patent US20160254487A1 "Permeation barrier system for substrates and devices and method of making the same." Figures 3, 4, and 5 show part of the text concerning the problem, field, and solution extracted from U.S. patent US20160254487A1, respectively. From these figures, our method is able to extract sentences of different aspects from the same patent. This affects the clustering results, as shown in Figure 1.

Table 3: Evaluation of clustering results using GMM, K-Means and Mean shift.

Clustering method	Evaluation measure	Structure tags			
		effects	field	problem	solution
GMM	MI	0.100	0.010	0.021	**0.000**
	NMI	0.153	0.014	0.035	**0.000**
	AMI	0.117	-0.003	0.017	**0.000**
	VM	0.153	0.014	0.035	**0.000**
K-Means	MI	0.040	0.016	0.055	**0.000**
	NMI	0.050	0.020	0.073	**0.000**
	AMI	0.014	0.006	0.059	**0.000**
	VM	0.050	0.021	0.073	**0.000**
Mean Shift	MI	0.131	0.054	0.049	**0.000**
	NMI	0.177	0.112	0.115	**0.000**
	AMI	0.039	0.026	0.040	**0.000**
	VM	0.177	0.112	0.115	**0.000**

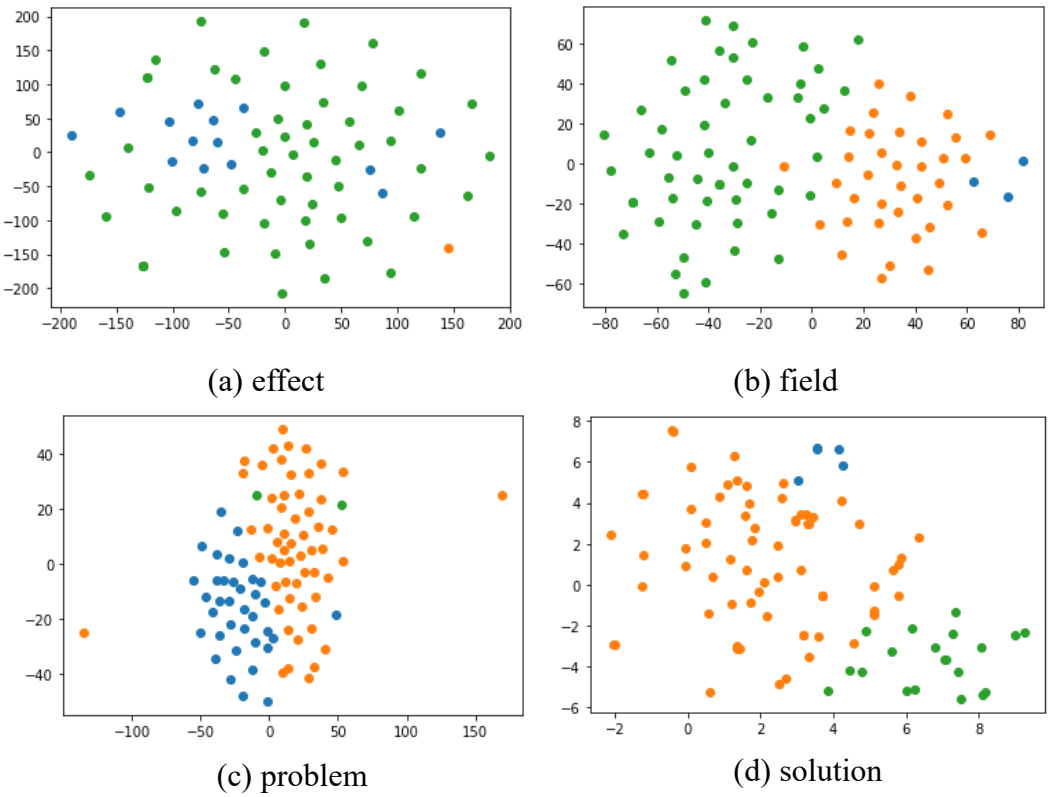

Figure 1: Clustering results using 94 U.S. patents in the high-barrier film field. Orange, blue and green dots indicate the three axes of analysis: film material, film-making material and film-making technology, respectively.

Disclosed is a novel moisture permeation barrier system for substrates and devices and method of making the same. The permeation barrier system includes two barrier layers. The first barrier layer is disposed over the substrate or an electronic device. The second barrier layer is then disposed over the first barrier layer. This system has relatively low permeability to moisture and is flexible. It may cover particles and provide moisture protection with a relatively small width edge seal.

Figure 2: Abstract of US patent US20160254487A1

Opto-electronic devices that make use of organic materials are becoming increasingly desirable for a number of reasons. OLEDs make use of thin organic films that emit light when voltage is applied across the device. OLEDs are becoming an increasingly interesting technology for use in applications such as flat panel displays, illumination, and backlighting.

Figure 3: Part of the sentences extracted from US patent US20160254487A1 regarding the problem

> The present invention relates to permeation barriers for devices such as organic light emitting diodes and other devices, and devices including the same. Several OLED materials and configurations are described in U.S. Pat. Nos. 5,844,363, 6,303,238, and 5,707,745, which are incorporated herein by reference in their entirety.

Figure 4: Part of the sentences extracted from US patent US20160254487A1 regarding the field

> The permeation barrier system includes two barrier layers. The first barrier layer is disposed over the substrate or an electronic device. The second barrier layer is then disposed over the first barrier layer. This system has relatively low permeability to moisture and is flexible. It may cover particles and provide moisture protection with a relatively small width edge seal.

Figure 5: Part of the sentences extracted from US20160254487A1 regarding the solution

6 Conclusion

In this study, we created a dataset containing 81,405 U.S. patents with structural tags. Using this dataset, we conducted an experiment to classify automatically sentences from the U.S. patents that are important for classifying each patent onto a technical analysis axis. The experimental results showed that one of our methods, patent-specific BERT, obtained an F-measure of 0.7426, which outperformed the others. We have analyzed the structure of U.S. patents using the proposed method and generated a technology trend map, which confirms the effectiveness of the proposed method.

References

[1] Vinodkumar Prabhakaran, William L. Hamilton, Dan McFarland, and Dan Jurafsky, "Predicting the rise and fall of scientific topics from trends in their rhetorical framing," In Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics, pp.1170–1180, 2016.

[2] Xiangci Li, Gully Burns, and Nanyun Peng, "Scientific discourse tagging for evidence extraction," In Proceedings of the 16th Conference of the European Chapter of the Association for Computational Linguistics, pp.2550–2562, 2021.

[3] Gully A Burns, Xiangci Li, and Nanyun Peng, "Building deep learning models for evidence classification from the open access biomedical literature," Database, 2019.

[4] Jinhyuk Lee, Wonjin Yoon, Sungdong Kim, Donghyeon Kim, Sunkyu Kim, Chan Ho So, and Jaewoo Kang, "BioBERT: a pre-trained biomedical language representation model for biomedical text mining," Bioinformatics, pp.1234–1240, 2020.

[5] Iz Beltagy, Arman Cohan, and Kyle Lo, "SciBERT: pretrained contextualized embeddings for scientific text," In Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing,

pp.3615–3620, 2019.

[6] Franck Dernoncourt and Ji Young Lee, "Pubmed 200k RCT: a dataset for sequential sentence classification in medical abstracts," In Proceedings of the 8th International Joint Conference on Natural Language Processing, pp.308–313, 2017.

[7] Gully APC Burns, Pradeep Dasigi, Anita de Waard, and Eduard H Hovy, "Automated detection of discourse segment and experimental types from the text of cancer pathway results sections," Database, 2016.

[8] Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova, "BERT: pre-training of deep bidirectional transformers for language understanding," arXiv preprint arXiv:1810.04805, 2017.

[9] Isao Goto, Ka Po Chow, Bin Lu, Eiichiro Sumita, and Benjamin K. Tsou, "Overview of the patent machine translation task at the NTCIR-10 workshop," In Proceedings of the 10th NTCIR Conference, pp.260–286, 2013.

[10] Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N. Gomez, Łukasz Kaiser, and Illia Polosukhin, "Attention is all you need," In Proceedings of Advances in Neural Information Processing Systems, pp.6000–6010, 2017.

[11] Myle Ott, Sergey Edunov, Alexei Baevski, Angela Fan, Sam Gross, Nathan Ng, David Grangier, and Michael Auli, "FAIRSEQ: a fast, extensible toolkit for sequence modeling," In Proceedings of North American Association for Computational Linguistics (NAACL): System Demonstrations, pp.48–53, 2019.

[12] Hamid Bekamiri, Daniel S. Hain, and Roman Jurowetzki, "PatentSBERTa: a deep NLP based hybrid model for patent distance and classification using augmented SBERT," arXiv preprint arXiv:2103.11933, 2021.

[13] Dorin Comaniciu and Peter Meer, "Mean shift: A robust approach toward feature space analysis," IEEE Transactions on Pattern Analysis and Machine Intelligence, Volume 24, Issue 5, pp. 603-619, 2002.

[14] Andrew Rosenberg and Julia Hirschberg, "V-Measure: A conditional entropy-based external cluster evaluation measure," In Proceedings of the 2007 Joint Conference on Empirical Methods in Natural Language Processing and Computational Natural Language Learning, pp. 410–420, 2007.

[15] Laurens van der Maaten and Geoffrey Hinton, "Visualizing High-Dimensional Data Using t-SNE," Journal of Machine Learning Research 9, pp. 2579-2605, 2008.