Seiya Maki [1], Satoshi Ohnishi [2], Minoru Fujii [1], Naohiro Goto [3]

Consideration of an Industrial Database Format for Resource Recycling and for Matching Recyclable Resources Using Between Industries

[1] National Institute for Environmental Studies, Japan. Social Division
[2] National Institute for Environmental Studies, Japan. Fukushima Regional Collaborative Research Center
[3] Toyo University.
†Corresponding author E-mail: maki.seiya@nies.go.jp
https://doi.org/10.65776/ep.20.2.2

Abstract

As one of the strategies to mitigate climate change and increase resource use efficiency, there is a need to improve energy recovery and recycling. Then, sector-coupling methods between industries are expected to improve resource recycling efficiency. This study considered the supply of extracted steam from waste treatment plants (WTP) to neighboring factories. We focused on the fact that the company names contained characteristic terms depending on the type of industry. The target factories of the steam supply were classified into industry types by the LSTM method based on the database of company names, and we estimated spatial steam demand using developed models However, they lack of spatial energy demand data could not validate this model result. From this situation, the process database of industries for matching recyclable resources and the list of company names for estimating the location of industries would need to be developed. Therefore, it is necessary to establish a beneficial system to provide data for validation. Based on the knowledge obtained from estimating the location of industries based on the company names and estimating the distribution of industrial steam demand, we developed a database arrangement model system. In this context, trial consideration of the industrial process with Resource Description Framework (RDF) formatting was conducted to enable the data collection and retrieval of process data through a knowledge base.

***Keywords**: Recyclable resource, Circular economy, Supply-Demand Matching, RDF, Semantic Web*

1. Introduction

In the field of waste management, improvement of energy recovery and acceleration of recycling are required in terms of global warming countermeasures and the improvement of resource efficiency. The sector coupling method is currently being focused on to improve the efficiency of resource recycling and the utilization of waste materials. However, t the supply and demand of recyclable resources, cannot be considered in terms of quantity and quality. Therefore, it is necessary to determine the spatial distributions of supply and demand.

Fujii et al. [1] explained steam that steam from waste treatment plants to some industries is more efficient regarding exergy, energy and economic benefits than electricity generation. This study focuses on the industrial steam demand for recyclable resource distribution.

Maki et al. [2] estimated steam demand using geographical data on industries, the steam supply potential of waste treatment plant, and the locations of waste generation sites in Aichi Prefecture, Japan. They found that the steam supply could be transported as far as 6 km from the generation point and that the waste treatment plants had a steam generation potential of -60%. Maki et. al. [3] improved on this study and analyzed industrial augmentation areas that should be prioritized as steam supply targets from an economic perspective.

In industrial ecology, Behera et al. [4] evaluated the effects of an exhaust heat utilization network and the heat supply from a waste incineration plant in an industrial park in Ulsan, South Korea. The authors also evaluated the business processes facilitating a targeted Eco-industrial park (EIP). Dou et al. [5] [6] [7] analyzed that included the industrial scenario based on a recent industrial location map.

Most studies that have focused on the geographic distribution of recyclable resources and attempted to improve the efficiency of resource circulation based on the geographic distribution of recyclable resources have focused on recycling. Wu et. al. [8] used GIS to estimate the future generation and management of construction and other demolition waste in Shenzhen City, China. Robinson et. al. [9] used GIS to study the appropriate arrangement of recycled aggregate use based on transportation networks and population density information. Nour Madi et. al. [10] used GIS to estimate the amount of demolition waste generated, including that generated during disasters, and considered recycling sites. Wang et. al. [11] used machine learning methods to estimate spatial building resource stocks in Japan. Wiedenhofer et. al. [12] estimated resource stocks in transportation infrastructure at national level. Bai et. al. [13] used a 4D-GIS to estimate building stocks and developed a bottom-up spatio-temporal database of carbon emissions from construction and demolition in Japan. These studies have estimated the amount and potential of waste and other resources generated; however, few have attempted to understand the existing demand locations for recyclable resources. Pauliuk et al. [14] noted that industrial ecology lacks generic structures and databases, necessitating the development of databases for components and analysis methodology systems.

In this situation, we focused our research on the existence of characteristic terms in each factory industry in the steam demand target. We developed an industry classification estimation model using the LSTM (Long Short-Term memory) model from the names of each factory listed in the Nationwide Factory Directory data. This text-mining method was used by Rizwan et al. [15], Rahimi et.al. [16], Lourentzou et al. [17], Han et.al. [18], Hasan et.al. [19], Luo et.al. [20] and many other studies that analyzed geographic information data associated with SNSs and estimated their geographic distribution. Several studies in this area were reviewed by Utomo et al. [21]. However, few studies have focused on company names or analyzed industries.

We developed an LSTM model that estimated the demand potential for steam at 200°C or lower. This model estimated industrial distribution of steam demand locations using text mining and statistical databases. The potential was estimated from the 1km mesh level, using the estimated industrial distribution based on steam demand unit data by industry. Because this research is limited, obtaining data on the demand potential of quantity and quality by industry for other recyclable resources is difficult. However, expanding the list of sample companies with their names and industrial types matched is necessary to improve accuracy. As data were not adequately collected, this study focused on developing a system to estimate industrial location. Improving the accuracy will be the focus of future studies. Data collection bias and data number problems were caused by the lack of a mechanism to collect and compile data in future studies.

2. Method

2.1 Factory type of steam demand target

In Factory name data statistics (Nationwide Factory Directory data [22] [23], the 28385 factories listed in the Kanto/Kansai region, where digital data is available, were targeted. The target industry sectors were selected based on the steam demand factory statistics (A comprehensive survey of energy consumption in industrial facilities,) [26]. However, the metal manufacturing industry, which has the largest sample size, is separate from the steel industry. The industry classifications are listed in **Table 1**.

Table.1 Steam demand target factory type [26]

① Foods (Processing)
② Foods (Ready meal)
③ Foods (Seasoning)
④ Foods (Bakery/Confectionery)
⑤ Beverages, tobacco & feed (Soft drink)
⑥ Beverages, tobacco & feed (Liquor)
⑦ Textile & Apparels
⑧ Pulp & paper products
⑨ Inorganic chemicals
⑩ Organic chemicals
⑪ Pharmaceutical
⑫ Oil and coal products
⑬ Plastic products
⑭ Rubber products
⑮ Ceramic, stone & clay products
⑯ Iron and Steel
⑰Metals
⑱Nonferrous Metals
⑲General machinery equipment
⑳Electronic parts, devices & electronic circuits
㉑Transportation equipment (automobile)
㉒Non-target

2.2. Industry Classification and Location Estimation Model from Company Name

Fig. 1 shows how to estimate the building type. It is difficult for existing factories to create such figures because of privacy issues and other reasons. Therefore, we created a figure based on a Japanese map using a public research institute affiliated with the author as the subject to illustrate a specific image of the analysis. From the textual information "研究所," it is possible to estimate that the building is an institute. Thus, specific terms are considered to characterize specific industries. We apply this characteristic to identify the type of industry for companies through text analysis. We aim to develop a system that estimates the spatial distribution of steam demand by integrating it into lists containing address data.

Fig. 1. Concept image
(ex., National Institute for Environmental Studies)

In this study, we developed an industry-classification model based on factory names using LSTM. The Deep Learning Toolbox and Text Analytics Toolbox in MATLAB R2022b were used for the analysis.

We developed an industry classification model using the LSTM model with word coding based on the text data of the target factory names. The analysis was conducted on a dataset in which 80% of the factories in each industry were randomly selected from the 22 targeted industries for the learning data. We examined two approaches to analysis: one using stop words and one that did not. Here, stop-words mean terms without meaning in text analysis, such as the Japanese symbol “㈱” meaning “ company limited.” The removal of such terms is known as stop-word processing.

For estimation at the national level, we developed a model to classify the manufacturing industry among corporations based on Phonebook data [24]. Phonebook data is a list of corporations registered in the Phonebook for each prefecture and contains data on nearly 10 million corporations nationwide. This database classified industrial categories as “manufacturing” or others, which we named “non-target.” We developed a model that combines manufacturing industry data for all prefectures and uses 5% of the manufacturing industries and the same number of non-target companies as the learning data.

The two models were then integrated. We developed a two-stage industry classification model using LSTM with word coding, as shown **Fig. 2**, and developed industry unit data for industries with steam demand below 200°C. Spatial industry and steam demand estimations were conducted at the national level. In this study, we estimate industry types based on company names. Based on these results, we developed a system to estimate steam demand by estimating the industry type and location. Therefore, we used the word-encoding method for text analysis, with a maximum of 10 words. After time-series data conversion, estimation was performed using a classification model set as a single-layer 80-perceptron LSTM. Because of limited training data and the inability to randomly sample industries within Japan, improving the accuracy of models may not always improve the accuracy of the spatial distribution estimation. Improving analytical accuracy, including problems with data collection, remains a challenge for future research.

Cross-validation analysis was also performed for each category to compare and evaluate the applicability of this study. We analyzed the trends in the sample size and correct rates based on cross-validation for each category.

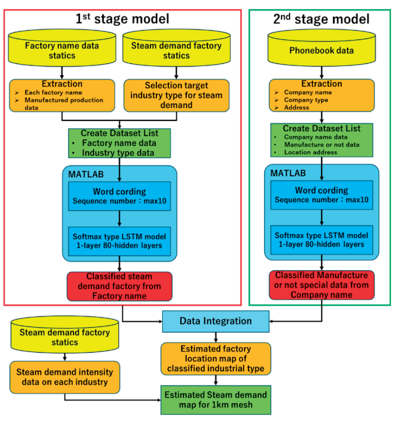

Fig. 2. Spatial Estimation Algorithm

3. Result

3.1. Industry Classification Model Results

We developed a model that could estimate industry types from company names with a correct response rate of about 70–80% for 22 industries (**Fig. 3.**).

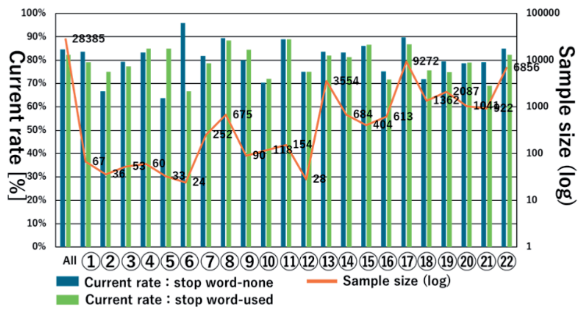

Fig 3. Results of the Industry Classification Model

Table 2 presents the cross-validation results. These results confirm that stop-word processing had no significant effect in cross-validation either. Table 2 shows that industries with larger sample sizes had higher correct rates.

The highest accuracy obtained through cross-validation was approximately 60%, indicating that further improvements in accuracy are possible by advancing the model. However, the lack of data on many industries is problematic.

Table 2. Result of cross-validation for LSTM

	validation sample	Current rate stop word-none	Current rate stop word-used
①	14	35.7%	28.6%
②	8	25.0%	37.5%
③	11	9.1%	9.1%
④	12	25.0%	25.0%
⑤	7	14.3%	28.6%
⑥	5	80.0%	60.0%
⑦	51	29.4%	39.2%
⑧	135	60.0%	58.5%
⑨	18	11.1%	22.2%
⑩	23	8.7%	8.7%
⑪	31	48.4%	54.8%
⑫	6	0.0%	0.0%
⑬	696	45.0%	52.9%
⑭	137	46.0%	45.3%
⑮	80	42.5%	47.5%
⑯	114	27.2%	23.7%
⑰	1663	63.1%	50.4%
⑱	262	19.8%	30.9%
⑲	401	24.9%	21.9%
⑳	199	22.6%	29.6%
㉑	173	28.9%	23.7%
㉒	1338	40.4%	36.6%
Total	5384	44.7%	41.4%

3.2. Results of the manufacturing classification model from the company name

Fig. 4 shows the results of the manufacturing classification. Each point represents a prefecture. The larger the number of factories, the higher the percentage of correct answers.

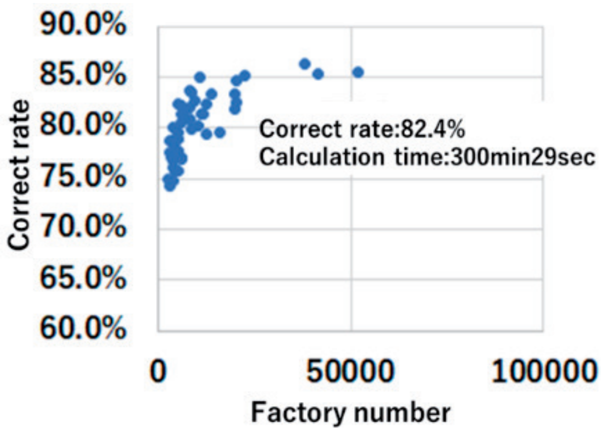

Fig. 4. Results of manufacturing classification model

3.3. Estimation of factory locations using a two-stage model

Based on these results, they estimate the spatial distribution of industries, as shown in **Fig. 5**. These figures are estimated for nationwide Japan, but because of the difficulty of description, Aichi and Yamaguchi prefectures are chosen for this report; each point in **Fig. 5** shows factory locations, and each color shows industrial types. If any point data were classified as non-manufacturing companies, they were eliminated from the map. Then, **Fig. 5** shows the estimated factory locations.

In Aichi Prefecture, many factories are distributed around the large population center of Nagoya, while in Yamaguchi Prefecture, many factories are located along the coast of the Seto-Nai Sea. It is thought that the results for Aichi Prefecture are affected by a large number of small factories located in areas with high population density. The Phonebook data lists factories without limiting their size, thus including small factories in the data. However, in some cases, factory names such as "rice cake (餅)" were classified as plastic product industries even though they were considered food processing industries by human thinking; this is thought to be caused by the lack of databases.

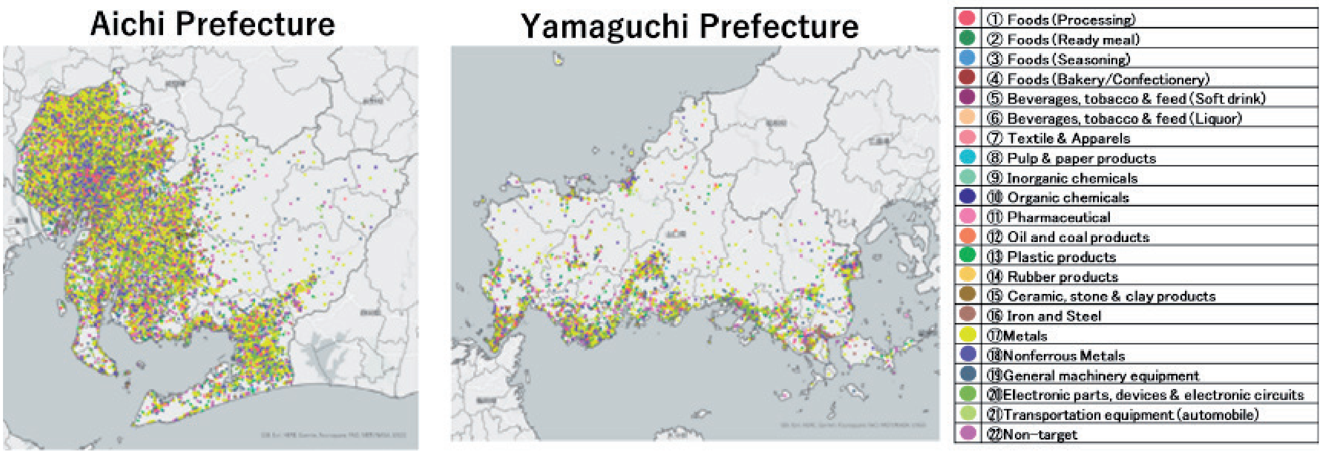

Fig. 5. Estimated results of industry distribution in Aichi and Yamaguchi Prefectures

3.4. Results of spatial estimation of steam demand

Fig. 6. was obtained in Aichi and Yamaguchi prefectures using steam demand unit data per office based on Reference [25]. Offices are typically tend to be located in highly populated areas. In Aichi Prefecture, it is estimated that a large amount of steam demand exists in the neighborhoods of cities, including inland areas. However, in Yamaguchi Prefecture, steam demand is concentrated near ports on the coast, and there are no large steam demand points in the inland area. In this context, we can develop a system to estimate the spatial steam demand based on company names.

However, this study did not sufficiently analyze the differences in the scale of the facilities. Using an average intensity value is considered to result in overestimating the heat demand; this was caused by using per-office intensity. Therefore, a future challenge is to proceed with estimates based on the building area and shipment value of manufactured goods.

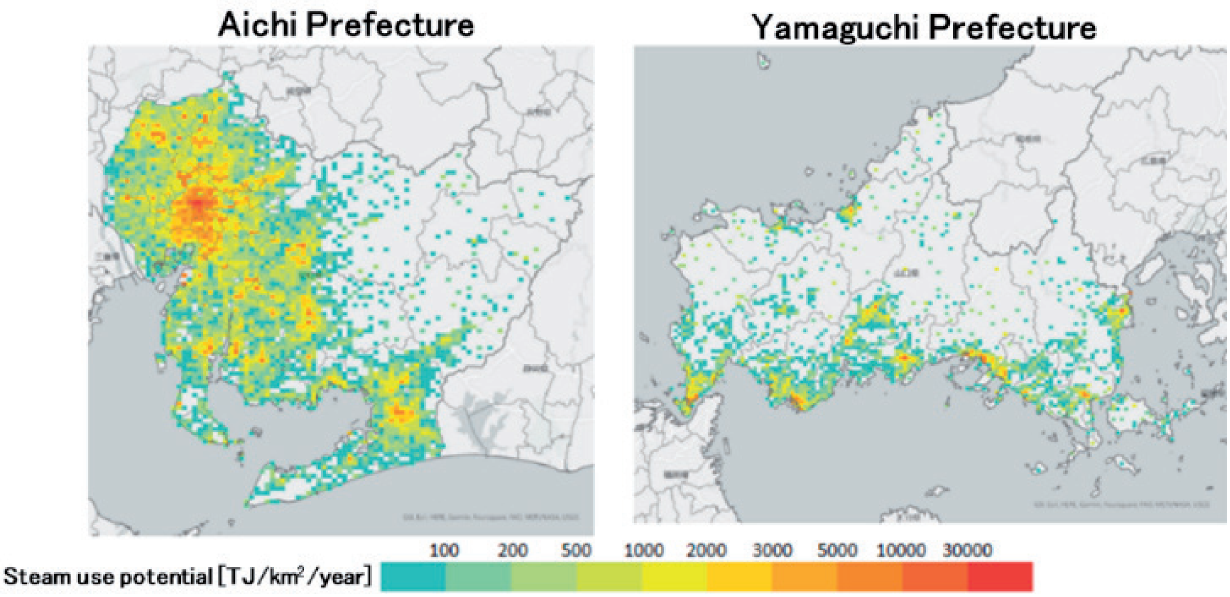

Fig. 6. Estimated results of the spatial distribution of steam demand in Aichi and Yamaguchi Prefectures

4. Conclusion

In this study, we considered the format structure of a database required for the supply-demand matching that utilizes recyclable resources.

We developed a model to classify industries based on factory names. This model was extended to Phonebook data, and a model was developed to classify manufacturing industries from corporations nationwide. By integrating the above two models, we estimated the location of factories that use steam and the estimated steam demand using the steam demand intensity of each industry.

As a result, we developed a model that can estimate the location of factories that are considered to have steam demand from nationwide corporations listed in the Phonebook data by text mining using the LSTM method. Using the steam demand intensity, we can estimate the amount of steam demand per 1 km mesh. The LSTM-based model achieved a certain level of accuracy, but further improvements were possible through the multilayering of the LSTM and hyperparameter tuning. However, there is a problem with the industry classification model in that the data from the main source does not comprehensively cover factories. However, some industries are not sufficiently classified. Although this study considers steam utilization, the model applies to other underutilized resources, if data are available. However, the lack of such data remains an issue.

5. Future works for the Necessity of data collection

Although it is possible to improve the accuracy rate by improving the LSTM model compared to the current data, the fundamental problems of bias in the limited data and the lack of data for validating the final factory location remain. Therefore, it is assumed that the advancement of classification models may not solve engineering problems because it is not beneficial to generalize data disclosure to the industrial side due to the inclusion of company secrets and technologies. However, the generalization and collection of industrial process data could increase the possibility of sector coupling among companies, thereby increasing the efficiency of recycling resource utilization. Therefore, it is necessary to consider a format that can promote sector coupling to utilize recyclable resources by increasing the number of samples and their types, as well as expanding the database.

In this situation, it is proposed to construct a data format in a knowledge base using an RDF [26], which is considered an easier way to build a relational database. RDF is also used for searches using search engines on the Internet. This technology is known as the Semantic Web. This technology can be used to estimate a knowledge base, such as the RDF described above, using the metadata on the HP; this makes it possible to search for recyclable-resource-use suggestions in a simplified way. **Fig. 7** shows an image of a connection based on several industrial processes using an RDF. In the Semantic Web field, methods for measuring the distance between knowledge bases such as RDF have been studied [27]. Because it is based on a semantic approach, the processes can be classified according to a linguistic database and thus can be presented without showing secret information about individual companies. Based on this technical background, we would like to explore expanding the sample data by developing data collection methods that utilize Semantic Web technologies.

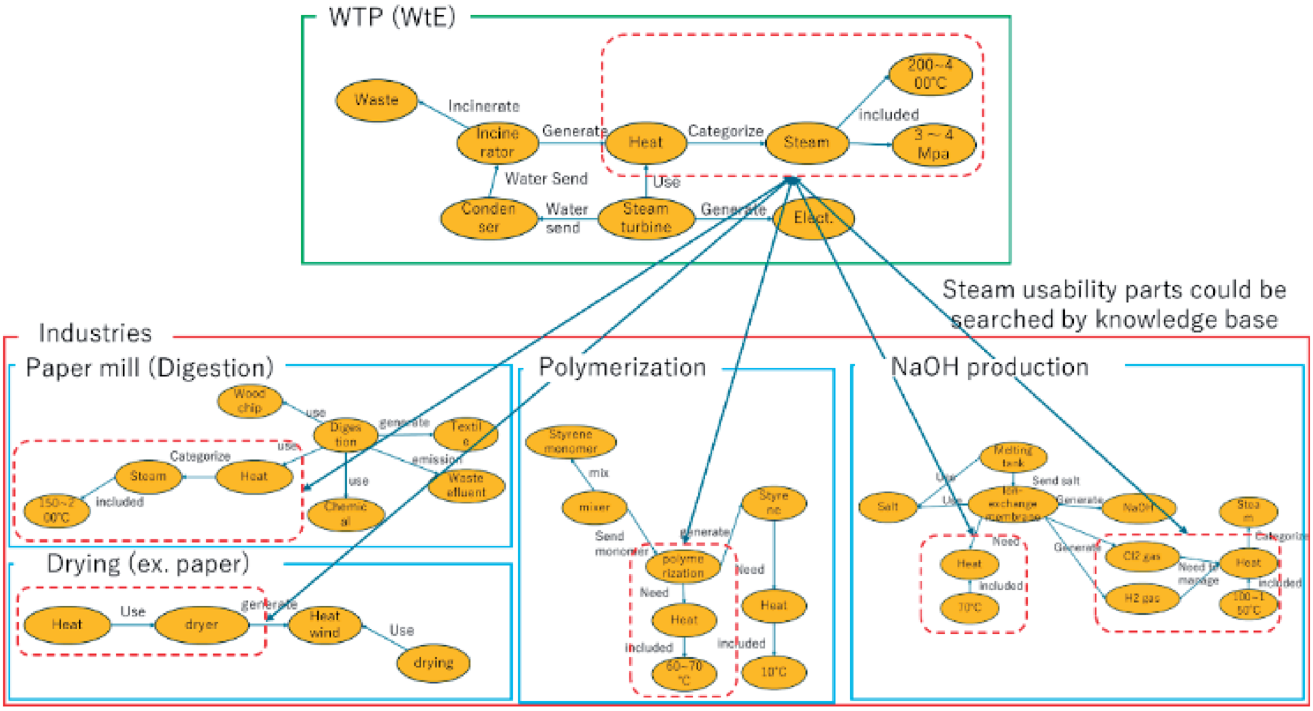

Fig. 7. Example of utilization of recyclable resources by the RDF description of industrial processes

6. Acknowledgments

This research project was supported by the following funding sources:

(i) JSPS KAKENHI Grant-in-Aid for Early-Career Scientists (21K14281)

(iii) Environmental Restoration and Conservation Agency of the Environment Research and Technology Development Fund (JPMEERF20201003)

(iiii) Environmental Restoration and Conservation Agency of the Environment Research and Technology Development Fund (JPMEERF20223001)

(iv) Environmental Restoration and Conservation Agency of the Environment Research and Technology Development Fund (JPMEERF20223C02)

(v) University Coalition for Carbon Neutrality. (AM0083-25)

7. References

[1] Fujii M., Dou Y., Sun L., Ohnishi S., Maki S., Dong H., Dong L., & Chandran R.. Contribution to a low-carbon society from improving exergy of waste-to-energy system by upgrading utilization of waste. Resources, Conservation and Recycling, 149, (2019) 586–594.

[2] Maki, S., Ohnishi, S., Fujii, M., Goto, N., & Lu, S. Technical and economic analysis of potential steam supply from waste treatment plants to industries in Aichi Prefecture, Japan. Optimization and Engineering, Vol.22, (2021) 1755–1782.

[3] Maki, S., Ohnishi, S., Fujii, M., Goto, N., & Lu, S. Using waste to supply steam for industry transition: Selection of target industries through economic evaluation and statistical analysis. Journal of Industrial Ecology, Vol.26, Issue4, (2022) 1475-1486.

[4] Behera, S.K., Kim, J.-H., Lee, S.Y., Suh, S., & Park, H.S. Evolution of "designed" industrial symbiosis networks in the Ulsan Eco-industrial Park: "research and development into business" as the enabling framework. Journal of Cleaner Production 29–30, (2012).103–112.

[5] Dou, Y., Ohnishi, S., Fujii, M., Togawa, T., Fujita, T., Tanikawa, H., & Dong, L. Feasibility of developing heat exchange network between incineration facilities and industries in cities: Case of Tokyo Metropolitan Area. Journal of Cleaner Production, 170(1), (2018).548–558.

[6] Dou, Y., Togawa, T., Dong, L., Fujii, M., Fujita, T., Ohnishi, S., & Kim, H. (2015). Development of a land use planning method for district heating system using waste heat: A case study in Fukushima, Japan. 43rd Annual Meeting of Environmental Systems Research, at Sapporo, (2015). 111–116.

[7] Dou, Y., Togawa, T., Dong, L., Fujii, M., Ohnishi, S., Tanikawa, H., & Fujita, T. Innovative planning and evaluation system for district heating using waste heat considering spatial configuration: A case in Fukushima, Japan. Resources, Conservation and Recycling, 128, (2018). 406–416.

[8] Huanyu Wu, Jiayuan Wang, Huabo Duan, Lei Ouyang, Wenke Huang, Jian Zuo. An innovative approach to managing demolition waste via GIS (geographic information system): a case study in Shenzhen city, China. Journal of Cleaner Production, Vol.112, (2016) 494-503.

[9] Gilpin R. Robinson Jr., Katherine E. Kapo. A GIS analysis of suitability for construction aggregate recycling sites using regional transportation network and population density features. Resources, Conservation and Recycling, Vol.42, Issue 4, (2004) 351-365.

[10] Nour Madi, Issam Srour. Managing emergency construction and demolition waste in Syria using GIS. Resources, Conservation and Recycling, Vol.141, (2019) 163-175.

[11] Lei Wang, Ruirui Zhang, Zhiwei Liu, Hiroaki Shirakawa, Hiroki Tanikawa. From expansion to efficiency: Machine learning-based forecasting of Japan's building material stocks under demographic declines. Science of The Total Environment, Vol.951, (2024) 175634.

[12] Dominik Wiedenhofer, André Baumgart, Sarah Matej , Doris Virág, Gerald Kalt, Maud Lanau, Danielle Densley Tingley, Zhiwei Liu, Jing Guo, Hiroki Tanikawa, Helmut Haberl. Mapping and modelling global mobility infrastructure stocks, material flows and their embodied greenhouse gas emissions. Journal of Cleaner Production,Vol.434, (2024) 139742.

[13] Yueyang Bai, Naho Yamashita, Jing Guo, Sota Nagata, Hiroaki Shirakawa, Hiroki Tanikawa. Dynamic analysis of carbon emissions from construction and demolition activities in Japan: Revealed by high-resolution 4D-GIS modeling. Journal of Environmental Management, Vol.360, (2024) 121099.

[14] Pauliuk, S., Heeren, N., Hasan, M. M., & Müller, D. B. A general data model for socioeconomic metabolism and its implementation in an industrial ecology data commons prototype. Journal of Industrial Ecology, 23(5), (2019) 1016-1027.

[15] Rizwan, Muhammad; Wan, Wanggen; Gwiazdzinski, Luc. Visualization, Spatiotemporal Patterns, and Directional Analysis of Urban Activities Using Geolocation Data Extracted from LBSN. ISPRS International Journal of Geo-Information, Vol.9, Issue 2, (2020) 137.

[16] Afshin Rahimi, Duy Vu, Trevor Cohn, Timothy Baldwin. Exploiting Text and Network Context for Geolocation of Social Media Users. https://doi.org/10.48550/arXiv.1506.04803 (Accessed Mar 26, 2025)

[17] Ismini Lourentzou, Alex Morales, ChengXiang Zhai. Text-based geolocation prediction of social media users with neural networks. 2017 IEEE International Conference on Big Data (Big Data), 2017.

[18] Bo Han, Paul Cook, Timothy Baldwin. Geolocation Prediction in Social Media Data by Finding Location Indicative Words. Proceedings of COLING 2012, (2012) 1045–1062.

[19] Samiul Hasan, Satish V. Ukkusuri, Xianyuan Zhan. Understanding Social Influence in Activity Location Choice and Lifestyle Patterns Using Geolocation Data from Social Media. Frontiers in ICT, Vol.3(10), (2016).

[20] Xiangyang Luo, Yaqiong Qiao, Chenliang Li, Jiangtao Ma, Yimin Liu. An overview of microblog user geolocation methods. Information Processing & Management, Vol.57, Issue 6, (2020) 102375.

[21] Muhammad Nur Yasir Utomo, Teguh Bharata Adji, Igi Ardiyanto. Geolocation prediction in social media data using text analysis: A review. 2018 International Conference on Information and Communications Technology (ICOIACT), (2018).

[22] The Nihon Sanki Shimbun: Nationwide Factory Directory data for Kanto, 2020(in Japanese)

[23] The Nihon Sanki Shimbun: Nationwide Factory Directory data for Kansai, 2020(in Japanese)

[24] EINS; Hello Page Phonebook CSV Data DVD Version Updated July 2021, 2021. (in Japanese)

[25] Vahid Ebrahimipour, Soumaya Yacout. Ontology Modeling in Physical Asset Integrity Management, (2015) Springer.

[26] Fuji Keizai: A comprehensive survey of energy consumption in industrial facilities 2017(in Japanese)

[27] Yong Jiang, Xinmin Wang, Hai-Tao Zheng. A semantic similarity measure based on information distance for ontology alignment. Information Sciences, Vol.278, Issue 10 (2014) 76-87.