Machine learning to improve waste management

Bei Gao (bgao@nuist.edu.cn)
Yunmeng Dai
Qin Bao
Yueqi Shang
Feier Han
Haiwen Yan
Shiyi Lu
Luofujie Guo
Ruidong Niu
Yifan Zhou
Yuen Jia
Jun Yin (jun.yin@nuist.edu.cn)

Short Report

Keywords: water management, waste recycling, clustering

Posted Date: May 11th, 2022

DOI: https://doi.org/10.21203/rs.3.rs-1644211/v1

Machine learning to improve waste management

Bei Gao[1,2*], Yunmeng Dai[3], Qin Bao[3], Yueqi Shang[3], Feier Han[3], Haiwen Yan[3], Shiyi Lu[3], Luofujie Guo[3], Ruidong Niu[3], Yifan Zhou[3], Yuen Jia[3], Jun Yin[2,3*]

[1]Department of Marine Environmental Science, Nanjing University of Information Science and Technology, Nanjing 210044, China
[2]Ecohydrology Lab, Nanjing University of Information Science and Technology, Nanjing 210044, China
[3]Department of Hydrology and Water Resources, Nanjing University of Information Science and Technology, Nanjing 210044, China

Corresponding to: BG (bgao@nuist.edu.cn) or J.Y. (jun.yin@nuist.edu.cn)

Abstract

While recycling facilities have been significantly upgraded in China, but the effectiveness of these facilities in improving waste management needs to be evaluated. Here, we conducted a nationwide survey by directly taking photographs of the inside of individual waste containers over 11 cities across China. We found that waste from recycling and non-recycling containers generally comprised similar materials. The corresponding waste features extracted by machine learning models tend to be well-mixed but clearly separated after removing the misplaced items, demonstrating an objective means for quantifying the accuracy of waste-sorting process. We therefore proposed the nationwide scale-up of this automated machine learning system, which along with additional incentive programs for better waste-sorting behaviors, may help improve waste management.

Main Text

With rapid urbanization and growing populations, global waste may increase by 70% in the next three decades, posing significant risks to human health, our living conditions, and natural ecosystems (*1*). There is a growing consensus that urgent actions must be taken to reduce the waste production and improve the management system for a sustainable planet. As one of the largest waste producers, China not only banned its import of plastic waste in 2018 but also implemented compulsory selective waste collection in 2019, showing great determination towards protecting the environment. The sales of selective waste containers remains at 2.5 million before 2018 but has quadrupled to 10 million in 2019 after the enforcement of selective waste collection (see Figure 1) (*2*), sharply transforming both the urban and suburban landscapes where newly installed waste containers can be easily spotted. It is estimated that the market value related to selective waste collection may reach as high as 16 billion U.S. dollar in China in 2022 (*3*).

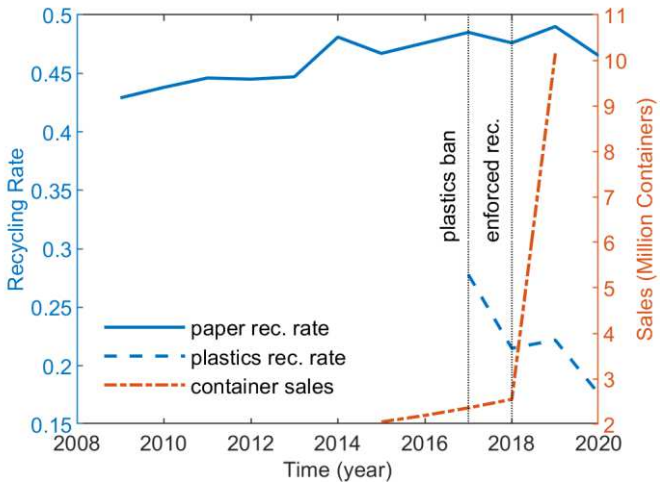

Figure 1. Comparisons of recycling rates and waste container sales. Recycling rates of paper and plastics were obtained from Statista; sales of selective waste containers were collected from business reports (see Methods).

While massive waste containers can be quickly installed, sorting waste into the right containers, a prerequisite for waste recycling, is one of the biggest waste management challenges. In fact, in contrast to the sharp increases of selective waste containers, the recycling rate of paper and cardboard basically remains unchanged and that of plastics even slightly decreases after 2018 (see Figure 1), suggesting that the newly installed selective waste containers may be not well used and requires further evaluation. Such evaluation provides direct feedback to the local communities, possibly contributing to the development of good waste sorting habits. As proven in Japan and South Korea, waste evaluation from sanitation workers along with strict rules forces citizens to reuse and recycling all possible materials, which has gradually become parts of their culture (*4*, *5*).

Towards this goal, here we conducted a nationwide survey by directly taking photographs of the inside of each selective waste container (see e.g., Fig S1) in the summer of 2020. These randomly surveyed containers are located in the public areas such as streets, campuses, parks, residential areas, shopping malls, and train/bus stations (see Methods), offering a glimpse of waste-sorting practices in typical public areas in China. To objectively evaluate whether the waste is sorted into the correct containers, we use pre-trained convolutional neural networks, which has been successfully used for classifying various images, including waste classification. Pre-trained network at the penultimate layer provides essential waste features, which were then fitted to their container types by using partial least squares regression (PLSR, see Methods). The score plots of the first few components allow us to check whether the waste features from the recycling and non-recycling containers are visually different, thus offering an objective tool for evaluating the accuracy of the sorting process (see Methods).

The score plot in Figure 2a shows that waste from recycling and non-recycling containers is essentially well-mixed with close centroids and largely overlapped

confidence intervals. The silhouette index, a measure of goodness of the classification, is close to zero, also indicating the waste is almost randomly classified. Therefore, we can conclude that the state-of-the-art machine learning cannot identify the visual differences of the waste from recycling and non-recycling containers. When carefully investigating these images, we found that dirty plastic bags or food scraps are often mixed in recycling containers and plastic/metal bottles are sometimes dropped in non-recycling containers. Once removing waste images with misplaced items, the new score plot in Figure 2b clearly shows that the waste is more appropriately classified and the mean silhouette index increases to 0.56, demonstrating that the patterns of well-sorted waste can be identified by machine learning classification.

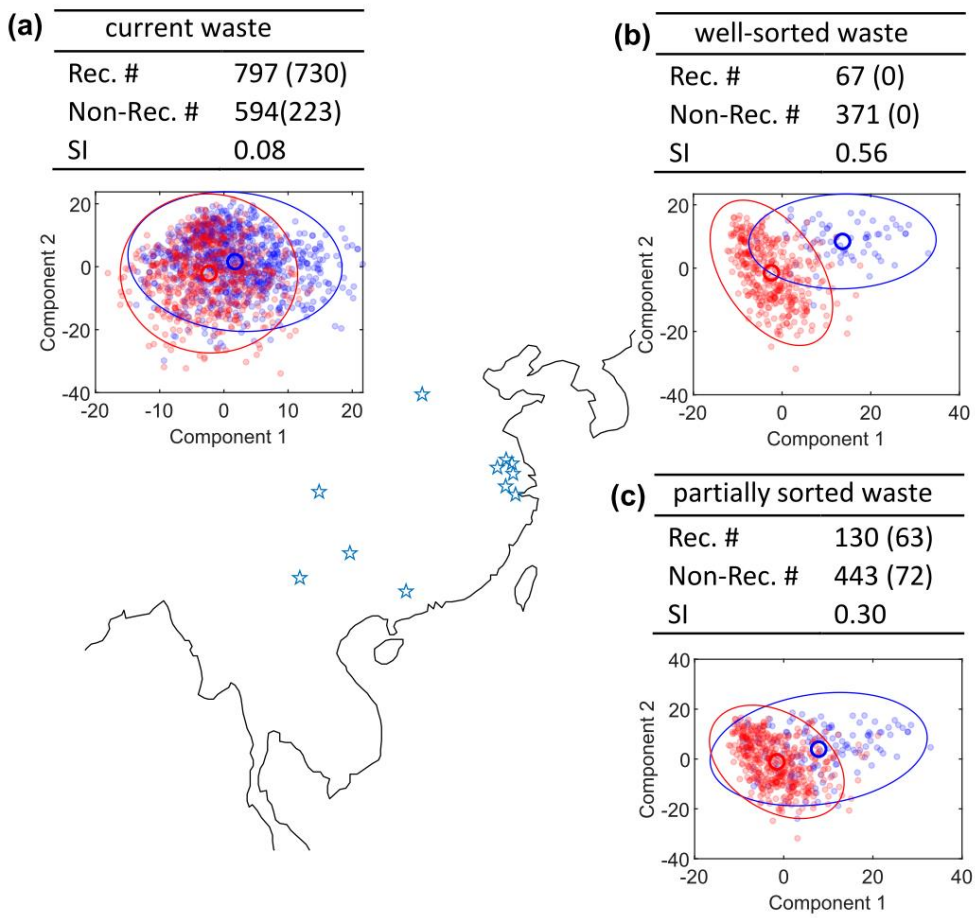

Figure 2. Waste survey locations and score plots. The stars are the waste survey locations over 11 cities across China (see Methods for the design of the survey). The insets are the corresponding score plots of waste features from (a) current waste, (b) well-sorted waste, and (c) partially-sorted waste (tables shows numbers of waste images followed by the numbers of misplaced items in the bracket and the silhouette index, SI); the blue and red dots indicate waste from recycling and non-recycling containers; the ellipses are the 95% confidence intervals.

The components of the well-sorted waste may be served a benchmark for

objectively evaluating waste-sorting efforts. When projecting onto these components, we found that waste in recycling and non-recycling containers from Yangtze Triangle Areas tend to be slightly better separated, consistent with the lowest misplacement rate identified from our visual check (see Fig S2). In general, the more misplaced items we have, the lower the silhouette index becomes (e.g., see Figure 2c). Therefore, machine learning classification and evaluation efficiently and objectively estimate to which degree waste is correctly sorted. This process, used to be a labor-intensive task performed by sanitation workers in many developed countries (*4*, *5*), now could be automated using machine learning models. Such automation makes possible a nation-wide waste-sorting feedback system and ultimately contributes to increase the waste recycling rate.

There are many ways to practically implement such an automated feedback system. For example, a specially designed smartphone app can be used by sanitation workers or residents to collect waste images and location information. Alternatively, waste images can be frequently taken by cameras installed above the waste containers. In fact, some smart containers have already installed cameras or other sensors (*6*). These images are then processed by machine learning and evaluation as proposed in this study. The evaluation results along with the waste locations are stored in the internet servers, allowing geographical comparison of the accuracy of waste-sorting process in different regions or communities. Reinforcement (e.g., rewarding communities with better waste-sorting behaviors) or peer pressure (e.g., sharing results on social media) may help citizens to develop good waste-sorting habits. Instead of linking to the numbers of the selective waste containers, government key performance indicators could be associated with the waste-sorting evaluation, thus forcing public sectors to pay more attentions to the real waste-recycling practices. As we have seen in the last few decades, China's top-down approach seems efficient at fixing environmental problems (*7*, *8*). We may expect that strict measures, including nationwide waste-sorting monitoring and evaluation system as well as citizen participation and oversight, may help improve waste management and protect our planet.

Competing interests

The authors declare no competing financial interests.

References

1. S. Kaza, L. Yao, P. Bhada-Tata, F. Van Woerden, *What a waste 2.0: a global snapshot of solid waste management to 2050* (World Bank Publications, 2018).
2. Guanyan, China's selective waste container market demand continues to grow in 2020. *Insight Info* (2021), (available at https://free.chinabaogao.com/fangzhi/202102/02225330N2021.html).
3. P. Chen, 100 billion market related to selective waste collection. *Bus. China* (2019).
4. C. K. Chrobog, South Korea: Cutting Back on Food Waste (2015), (available at https://pulitzercenter.org/stories/south-korea-cutting-back-food-waste).

5. P. McInerney, Recycling in Japan – Confusing yet Effective (2015), (available at https://contentedtraveller.com/recycling-japan-confusing-yet-effective/).

6. O. Ali, Managing Waste with Smart Bins (2022), (available at https://www.azosensors.com/article.aspx?ArticleID=2542).

7. X. Lu, S. Chen, C. P. Nielsen, C. Zhang, J. Li, H. Xu, Y. Wu, S. Wang, F. Song, C. Wei, K. He, M. B. McElroy, J. Hao, Combined solar power and storage as cost-competitive and grid-compatible supply for China's future carbon-neutral electricity system. *Proc. Natl. Acad. Sci.* **118**, e2103471118 (2021).

8. A. Viña, W. J. McConnell, H. Yang, Z. Xu, J. Liu, Effects of conservation policy on China's forest recovery. *Sci. Adv.* **2**, e1500965 (2016).

9. M. Tan, Q. Le, in *International conference on machine learning* (PMLR, 2019), pp. 6105–6114.

Supplementary Materials for

Machine learning to improve waste management

Bei Gao[1,2*], Yunmeng Dai[3], Qin Bao[3], Yueqi Shang[3], Feier Han[3], Haiwen Yan[3], Shiyi Lu[3], Luofujie Guo[3], Ruidong Niu[3], Yifan Zhou[3], Yuen Jia[3], Jun Yin[2,3*]

[1]Department of Marine Environmental Science, Nanjing University of Information Science and Technology, Nanjing 210044, China
[2]Ecohydrology Lab, Nanjing University of Information Science and Technology, Nanjing 210044, China
[3]Department of Hydrology and Water Resources, Nanjing University of Information Science and Technology, Nanjing 210044, China

Corresponding to: BG (bgao@nuist.edu.cn) or J.Y. (jun.yin@nuist.edu.cn)

This PDF file includes:

Methods
Fig. S1-S2

1. Methods

Waste Survey

In summer 2020, a nationwide waste survey is conducted by randomly selected undergraduate students from Nanjing University of Information Science and Technology, who were on their summer vacation either on road trips or at their hometowns. During their visits in public areas (e.g., streets, campuses, parks, residential areas, shopping malls, and train/bus stations), they took photographs inside the recycling and non-recycling containers using their smartphones. Overall, there are 1391 waste images labeled by their corresponding container types and the global position system information. The surveyed sites are presented in Figure 2 and an interactive map for these sites is available at shorturl.at/cwDHQ. All waste images are archived at figshare (doi.org/10.6084/m9.figshare.19705702).

Machine learning Classification and Evaluation

We used pre-trained network and dimension reduction to perform waste classification, and the goodness of clustering evaluates the waste-sorting efforts (see Fig S1 in supplementary materials). Specifically, we extracted image features by using the state-of-the-art EfficientNet, which is one of the most efficient models for image classification (*9*). Note that results in the main text are based on B4 variants of EfficientNet while B0-B7 variants or other popular pre-trained network show similar results. As in transfer learning, the penultimate layer of EfficientNetB4 provides the visual features of the waste images, a $n \times m$ matrix X, where n is the number of images (i.e., 1391) and m is the number of the extracted waste features (i.e., 1536 for EfficientNetB4).

Given the large number of waste features, we performed partial least regression (PLSR) against the corresponding container types, a $n \times 1$ binary vector Y, in which 1 is for the recycling container and 0 for the non-recycling container. Similar to principal component analysis, this regression allows us to decompose X as

$$X = TP^T + E \,. \tag{1}$$

where T is a $n \times p$ score matrix, P is a $m \times p$ loading matrix, p is the number of reserved components, and E is the error term. Such decomposition maximizes the covariance between T and Y, allowing us to explore the major differences between the waste from the recycling and non-recycling containers.

If the waste is well-mixed, it is impossible for PLSR to identify the visual differences in different containers and the corresponding score plots of T classified by vector Y is also well mixed as shown in Figure 2a in the main text. If the waste is well-sorted, the corresponding score plots for images from recycling and non-recycling containers are well separated as shown in Figure 2b in the main text. The goodness of clustering (e.g., mean silhouette index or other similar metrics) quantifies to which degree the waste is correctly sorted. To facilitate this procedure, we do not need to perform PLSR every time, but project any new waste features onto the principal components of the well-sorted waste. The new score plots will be better clustered for waste with less misplaced items (e.g., Figure 2c), thus expediting the

evaluation of the waste-sorting efforts.

2. Supplementary Figures

Fig S1. An example of waste field survey. Waste images were taken inside the containers and labeled as the recyclable or non-recyclable with GPS information.

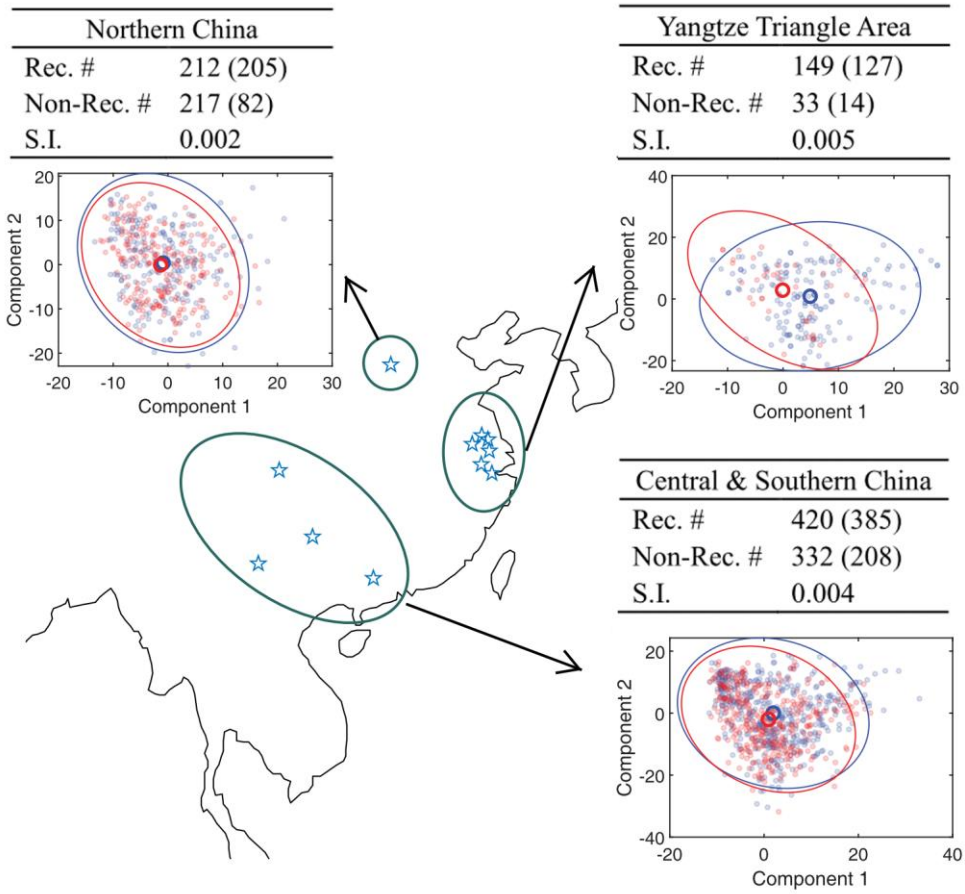

Fig S2. As in Figure 2 in the main text but for different regions. The waste features were projected onto the partial-least square regression components of Figure 2b in the main text. Twenty-eight waste images without GPS information were excluded in this regional analysis.