Waste Item Classification

Mu-Ruei Tseng, **Tara Poteat**, **Ankit Verma**

Introduction

Managing waste efficiently is a significant challenge. In 2015, 262 million tons of solid waste were processed in the United States alone. 91 million tons of this waste was recycled[1]. Clearly, recycling plays a large role in waste management. The most prominent recycling system in use currently is the Single Stream Recycling System, which involves manual item sorting by humans[2]. However, classifying waste manually is both time inefficient and space inefficient. This leads us to believe that machines will soon have a greater responsibility in classifying waste items. Through this project, we will develop a model that leverages machine learning algorithms to identify an item's 'waste-type' based on an input image.

Methods

The first step in our expected methods involves splitting our pre-classified dataset into training and test data. According to a research paper on splitting datasets by Dobbin and Simon, splitting datasets into ⅔ training data and ⅓ test data is a rule of thumb that works well in most scenarios[3]. The authors also mention, however, that even with a ⅔ to ⅓ split, data samples in each split can end up being unreliable due to chance[3]. So, in order to mitigate the possibility of unreliable data splits, we plan on implementing resampling of our data to find the optimal training-testing split.

Once our data is set-up and ready to go, our method involves training and testing of our dataset with three algorithms generally considered to be advantageous for multi-class classification problems. We will choose the optimal performing algorithm for our use-case after comparing and evaluating the performance of the following three supervised learning algorithms:

- Support Vector Machine
- Artificial Neural Network
- Decision Tree

Results

After being trained on training data, our model should accurately classify testing images of waste items into categories of either 'cardboard', 'plastic', 'glass', 'metal', 'paper', or 'trash'. We define successful classification as >95% of correctness. In case our data ends up being unbalanced, we will define successful classification as >0.95 'area under ROC'. To optimize our models chance of successful classification, we will evaluate three different machine learning methods for multi-class classification, and finalize our model with the evaluated model that performs best under our definition of successful classification.

Discussion

If further developed, the waste item classification model that results from our project can prove to be of value to waste management facilities that engage in automated sorting. Our project is limited in scope since we are identifying images of a confined format containing singular items. For a real-world use case, analyzing a photo for each and every trash item would likely be slow and cost-ineffective. We can further improve upon our proposed model by training it to be able to classify types of waste in pictures containing many items at once. This development would significantly shorten the amount of time it would take to sort the items, and prove more cost-effective to waste management facilities.

References

1. US EPA. (2018, June).*Advancing Sustainable Materials Management: 2015 Fact Sheet.* Retrieved from: https://www.epa.gov/sites/production/files/2018-07/documents/2015_smm_msw_factsheet_07242018_fnl_508_002.pdf
2. LeBlanc, R. (2019, June 25). Single Stream Recycling Offers Benefits, Creates Challenges. Retrieved October 1, 2019, from https://www.thebalancesmb.com/an-overview-of-single-stream-recycling-2877728.
3. Dobbin, K. K., & Simon, R. M. (2011, April 8). Optimally splitting cases for training and testing high dimensional classifiers. Retrieved October 1, 2019, from https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3090739/.