DOI: 10.37943/AITU.2021.52.74.005

DOI: 10.37943/AITU.2021.52.74.005

D. Chinassylov
BSc student of Software Engineering, Department of Computer Engineering
d.chinasylov@astanait.edu.kz, orcid.org/0000-0003-0944-3476
Astana IT University, Kazakhstan

A. Kozhamseitova
BSc student of Software Engineering, Department of Computer Engineering
a.kozhamseitova@astanait.edu.kz, orcid.org/0000-0002-2678-1795
Astana IT University, Kazakhstan

M. Kalen
BSc student of Software Engineering, Department of Computer Engineering
m.kalen@astanait.edu.kz, orcid.org/0000-0002-2377-2391
Astana IT University, Kazakhstan

R. Omirgaliyev
MSc of Electrical Engineering, Department of Computer Engineering
Ruslan.omirgaliyev@astanait.edu.kz, orcid.org/0000-0003-3834-2359
Astana IT University, Kazakhstan

APPLICATION INFORMATION MODELING AND MACHINE LEARNING ALGORITHM FOR CLASSIFICATION OF WASTE USING SUPPORT VECTOR MACHINE

Abstract: The ecological state of the world is deteriorating for the worse every year. One of the main problems is inadequate waste disposal and inadequate sorting by waste type, which has led to inadequate treatment of bulk waste in landfills throughout the world. The issue of improper disposal of municipal solid waste (MSW) in Kazakhstan has been raised since 2013, to solve this problem, the first President of the Republic of Kazakhstan, Nursultan Abishevich Nazarbayev, issued a decree on the transition to a green economy. Under the leadership of the Ministry of Energy, it was planned to reduce the amount of inappropriate waste by 40% in the territory of Kazakhstan by 2030. There are a lot of problems in India like inadequate waste collection, transport, treatment, and disposal. Poorly recyclable garbage has a global impact, fouling oceans, obstructing sewers, and creating flooding, transferring infections, increasing respiratory problems due to burning, injuring animals that inadvertently consume waste, and affecting economic development. To classify garbage, researchers utilized a combination of mixed modeling and machine learning techniques. Using machine learning technology, the data obtained can be used to classify and redistribute garbage for any sector around the world.

Keywords: image classification, support vector machines, principal component analysis

Introduction

The relevance of the work is associated with the need to develop and implement methods for modeling and analyzing the processes of waste classification, which can lead to reducing

Scientific Journal of Astana IT University
ISSN (P): 2707-9031
ISSN (E): 2707-904X

toxic waste ending in landfills. It is worth noting that waste recycling all over the world is quite a lucrative business. There are small factories in Kazakhstan that "bring garbage back to life". However, their capacity is not enough to reduce the filling of landfills. One of the unsolved problems of such plants is the lack of sufficient automation. Fortunately, the computer-aided power supply control system allows these companies to sort waste as well as reduce the health risks associated with human labor. Machine analysis allows classifying waste by minimal costs such as human resources and waste redistribution. This technology requires the use of modern information technologies, intelligent decision support systems, and waste classification management. The automated waste classification management system allows for a complete analysis of waste sorting, taking into account all the features of the waste. Therefore, the article used machine learning methods to classify waste. Support Vector Machines classification model was used to determine whether an object is recyclable or not and compared the training and testing results. Site-specific classification monitoring involves the process of continually comparing waste data with other features such as organic and recyclable waste. Support Vector Machines classification model was used to determine whether an object is recyclable or not and compared the training and testing results.

Aim:

The purpose of this research is to automate the process of classification of organic and non-organic wastes by using machine learning algorithm, namely SVM (Support Vector Machines).

Literature review

The search for efficient classification algorithms began with the inception of machine learning to this day. People from all over the world are developing algorithms for greater classification accuracy. The automatic compost detection system was developed by Google TensorFlow in 2016. One of the significant disadvantages of this system was that it could only detect compost materials. Led by Alex Krizhevsky, AlexNet [1] was proposed, which has made very good advances in the field of automatic image classification. Since that point, ultra-precise neural networks have been proposed that are likely to consequently distinguish and classify targets. But a new method for classification was proposed by Noushin Karimian et al., where it can classify three metals and can construct the most efficient classifier at the time. Also, Yusoff. S. H. researched and created a system that can automatically classify household metal waste. Zeng et al. We studied a method for automatically detecting the spread of waste over large areas using hyperspectral data, which made it possible in the future to improve the system for classifying waste. Hence, a modern hyperspectral image classification network has been created, which has demonstrated itself well within the automatic detection of waste over large regions. Then, there was produced a hyperspectral imaging system for collecting tests of different types of waste and pre-processed the samples for noise reduction and correction, which made it possible to obtain more accurate classification results. Subsequently, this technology was used to further classify waste. SeokBeom Roh et al. created a hybrid technology to produce a neural network classifier with a radial basis function that can efficiently recycle waste.

Kennedy et al. employed the Visual Geometry Group 19 (VGG-19) as the primary demonstration of exchange learning; the categorization accuracy of waste photos was 88.4%, utilizing VGG-19's capacity to extract features [2]. Adeeji used the support vector machines (SVMs [4]) to categorize the convolution neural network model created by the 50-layer residual network preprocessing (ResNet-50 [3]) as the extractor. Using the SVM technique, 87% accuracy was attained on the garbage image dataset. Chen Zhihong et al. proposed a garbage grab system based on computer vision and an automatic sorting robot. Recognizing objects and estimating

DOI: 10.37943/AITU.2021.52.74.005

attitudes were achieved using the Region Proposal Network (RPN) and VGG-16 models [5]. The model was created using MobileNet by Stephen L. and others. For the Imagenet large-scale visual identification test, transferred learning achieved an accuracy of 87 percent [6]. Stephen L et al. improved and quantified the model in order to successfully apply to mobile devices, and the accuracy reached 89.39 percent, which is a very good result. Ruiz V. et al. examined different deep learning systems in the automatic classification of waste types and exploited the benefits of traditional deep learning models. The best combination of the ResNet model yielded 88.6 percent accuracy on garbage photos. Costa and colleagues looked at various types of neural networks and categorized the trash photos into four groups. The accuracies of the Support Vector Machines (SVM), K Nearest Neighbors (KNN), and Random Forest (RF) pretraining model techniques were 85.0%, 80.0%, and 88.0%, respectively, across the diverse neural networks [7].

Traditional machine learning techniques necessitate a large amount of calibration training data, which necessitates a large number of human and fabric assets. Traditional machine learning algorithms such as Multi-Layer Perceptron (MLP), K-Nearest Neighbors (KNN), and Random Forest (RF) perform a huge number of calculations and may not match the data and adjust the samples well. Therefore, building on the previous statements, the traditional Machine learning algorithms are not suitable for waste classification. Among the neural network strategies for classifying waste, most of them utilize classical convolutional neural networks for fine-tuning or pre-training on huge datasets. The pre-workout approach and fine-tuning, on the other hand, contains many parameters, and fine-tuning of small datasets can result in overestimation or underestimation. The use of pre-training models and fine-tuning on tiny datasets may not be the optimal technique to fit data, according to the research. Furthermore, the waste classification literature cited above is insufficient. SVM is 87 percent better at effectively classifying junk and implementing the aforementioned tasks for these activities [1].

All materials wasted, whether recycled or disposed of in a landfill, are included in waste generation. The impact of new projects on the local waste stream can be estimated using waste generation rates for residential and commercial operations. As a result, in order to achieve efficient and orderly solid waste management, the essential components and linkages involved must be defined, modified for data uniformity, and properly understood. In a densely populated area, indiscriminate solid waste disposal and collection system failure would quickly result in health issues. In such a case, it is critical to settle the problem as quickly as feasible. And to do so, you'll need to automate all of the plant's trash sorting and recycling activities. Waste management and efficient sorting have long been seen as critical components of ecologically sustainable growth. Recycling and reusing discarded products are critical for society to reduce waste accumulation. To increase recycling and reduce environmental effects, efficient selective sorting is frequently used [6]. This issue should be addressed in particular in emerging countries, where the waste management is a major barrier to urbanization and economic growth.

Waste classification requires certain calculations, for which functions such as macro, micro, and weighted average calculations have been used. Micro and macro averages (for any metric) will calculate somewhat different things, and their interpretations will differ as a result. The macro average calculates the metric separately for each class before taking the average (therefore treating all classes equally), but the micro average aggregates all class contributions to calculate the average metric. When setting up a multiclass waste classification, it is preferable to use the macro average if you suspect that there may be an imbalance of classes (i.e., you may have many more examples of one class than other classes).

Scientific Journal of Astana IT University
ISSN (P): 2707-9031
ISSN (E): 2707-904X

To demonstrate the reason, take precision as an example

$$Pr = \frac{TP}{(TP+FP)} \quad (1)$$

Assume that you have a *One-vs-All* multi-class classification system with four classes and the following numbers when tested:

		Actual Class			
		A	B	C	D
Predicted Class	A	10 (TP)	0 (FP)	1 (FP)	0 (FP)
	B	90 (FP)	1 (TP)	0 (FP)	0 (FP)
	C		0 (FP)	1 (TP)	1 (FP)
	D		1 (TP)	0 (FP)	1 (TP)

Class A: 10 TP and 90 FP
Class B: 1 TP and 1 FP
Class C: 1 TP and 1 FP
Class D: 1 TP and 1 FP
We get values such as $Pr_B=Pr_C=Pr_D=0.3$ whereas $Pr_A=0.1$
A macro-average will be computed: $Pr = 0.1+0.54+0.5+0.5=0.4$
A micro-average will be calculated: $Pr = 1+10+1+12+100+2+2=0.123$
Here, TP means True positives (namely, those examples, which were predicted correctly), FP – False positives (those examples with wrong prediction)

These are two very distinct precision levels. Intuitively, the «excellent» precision (0.5) of classes A, C, and D contributes to a «decent» overall precision in the macro-average (0.4). While this is technically correct (the average precision across classes is 0.4), it is a little misleading because a huge number of samples are incorrectly categorized. Despite accounting for 94.3 percent of your test data, these examples mostly correspond to class B, therefore they only contribute 1/4 to the average. This class imbalance will be properly captured by the micro-average, lowering the overall precision average to 0.123 (closer to the precision of the dominant class B (0.1)).

It may be more convenient to compute class averages first and then macro-average them for computational reasons. There are numerous approaches to deal with class imbalance if it is recognized as a problem. One is to include not only the macro-average but also its standard deviation in the report (for 3 or more classes). Another option is to calculate a weighted macro-average, in which each class's contribution to the average is weighted by the number of examples available for that class. As a result of the above scenario, we get:

$$Pr_{macro\text{-}mean}=0.250.5+0.250.1+0.250.5+0.250.5=0.4$$
$$Pr_{macro\text{-}stdev}=0.171$$
$$Pr_{macro\text{-}weighted}=0.01880.5+0.9230.1+0.02890.5+0.02890.5=0.142$$

The substantial standard deviation (0.171) indicates that the 0.4 average does not reflect uniform precision across classes; however, it may be simpler to compute the weighted macro-average, which is essentially the same as the micro-average. As a result, we can classify things as follows:

DOI: 10.37943/AITU.2021.52.74.005

Classification report for train-set:

	precision	recall	f1-score	support
0	0.97	0.96	0.96	7505
1	0.95	0.96	0.95	6032
accuracy			0.96	13537
macro avg	0.96	0.96	0.96	13537
weighted avg	0.96	0.96	0.96	13537

Classification report for test-set:

	precision	recall	f1-score	support
0	0.83	0.80	0.81	2487
1	0.76	0.79	0.78	2026
accuracy			0.80	4513
macro avg	0.79	0.80	0.79	4513
weighted avg	0.80	0.80	0.80	4513

Classification report for CV:

	precision	recall	f1-score	support
0	0.82	0.79	0.81	2572
1	0.74	0.77	0.76	1941
accuracy			0.78	4513
macro avg	0.78	0.78	0.78	4513
weighted avg	0.79	0.78	0.79	4513

Fig. 1. Classification report for garbage dataset

Review of Support Vector Machine algorithm

Support Vector Machines (SVMs) are machine learning algorithms that are used for classification and regression purposes. SVMs are one of the powerful machine learning algorithms for regression, classification, and outlier detection purposes. An SVM classifier creates a model that assigns new data points to one of the given categories. Therefore, it can be regarded as a non-probabilistic binary linear classifier [8].

A hyperplane is a decision boundary that separates between a given set of data points having different class labels. The SVM classifier separates data points using a hyperplane with the maximum amount of margin. This hyperplane is known as the maximum margin hyperplane and the linear classifier it defines is known as the maximum margin classifier.

SVM produces an n-1-dimensional separation hyperplane to separate the two classes, and the distance between the hyperplane and the data points on each side is maximized. The goal of SVM is to determine the optimal hyperplane for separating the two classes [9].

Data are represented as

$$(x_1, y_1), \ldots, (x_n, y_n) \tag{2}$$

where y_i is either 1 or -1, indicating to which class x_i belongs. Each x_i is p-dimensional vector representing all of the characteristic values (variables) of x_i. The hyperplane that best separates the group of x_i vectors where $y_i = 1$ from the group of vectors where y_i = -1 is

$$x - b = 0 \tag{3}$$

Where is the hyperplane's normal vector, and b is the hyperplane's offset from the root. If the data points are linearly separable, the hard margin can be represented as

Scientific Journal of Astana IT University
ISSN (P): 2707-9031
ISSN (E): 2707-904X

$$x - b = 1$$

and

$$x - b = -1 \quad (4)$$

All these statements are drawn in Figure 2. It shows a maximum margin separation for linearly separable data. The samples that fall on the margin are known as the support vectors. SVMs find an optimal hyperplane to divide various classes with respect to X_1 and X_2 features, as shown in the graph. Finding the line that best separates two classes is the most basic example. The term 'best separates' refers to maximizing the profit margin. The placement of new observations in relation to the decision boundary is then used to classify them.

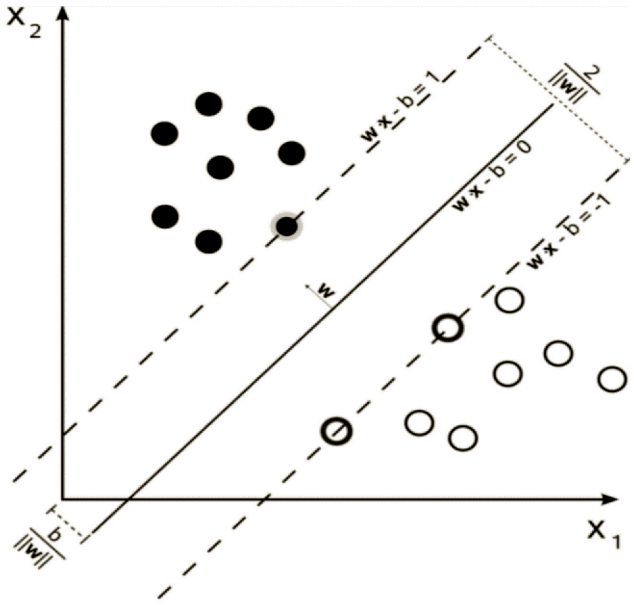

Fig. 2. Maximum margin hyperplane

Table 1. Advantages and Disadvantages of SVM model

Model type	Advantages	Disadvantages	Reference
SVM	1. SVMs work relatively well, when there is clear margin of division between two or more classes.	1. SVM algorithm is not appropriate for large data sets.	[10]
	2. SVM works more effectively in high dimensional spaces.	2. SVM does not work well when target classes are overlapping, the data set has more noise.	[10]
	3. SVM is effective in cases where the number of dimensions is greater than the number of samples.	3. SVM algorithm will underperform in such situations when the number of features for each data point surpass the number of training data samples.	[10]
	4. The SVM algorithm uses relatively a small amount of memory.	4. In SMV algorithm, there is no probabilistic explanation for the classification, because the support vector classifier works by putting data points above and below the classifying hyperplane.	[10]

DOI: 10.37943/AITU.2021.52.74.005

Results and Discussion

We are going to use a dataset containing 22564 images as input features and classify them into organic waste or recyclable waste using the Support Vector Machine algorithm and do some diagnoses. Figure 3 shows that the general process of the waste classification method consists of 4 steps:

Firstly, to save time, it can be useful to convert our dataset to an array form, so that our model can train on it.

Secondly, we are splitting the dataset into training, test and cross-validation sets. Then, we are performing dimensionality reduction using PCA algorithm on a training set to plot the dataset on the screen and speed up SVM algorithm execution time. Because images from the dataset contains lots of pixels, which means lots of features, it was necessary to use principal component analysis (PCA) in order to reduce the size of input parameters.

For the classification part, firstly, we train the SVM Classification model. Then, make a prediction on test and cross validation sets. As shown in Table 2, test and cross validation sets were predicted with 88% and 86% respectively. According to these statements, we can say that we are able to classify waste with approximately 88% accuracy, which is not a bad result. Once we are done with dimensionality reduction and classification, we will now plot our classification results of the training set. A dimensional reduction to 2-D helps us to plot results in order to make sense of how the SVM algorithm classified a dataset into two categories. Figure 4 shows classified objects on the graph, where "0" (blue dots) is organic waste and "1" (yellow dots) is recyclable waste, which is divided by the hyperplane (red line). SVM classifiers work by drawing a straight line between two classes, as previously stated. It means that all of the data points on one side of the line will be assigned to one category, while the data points on the other side will be assigned to a different category.

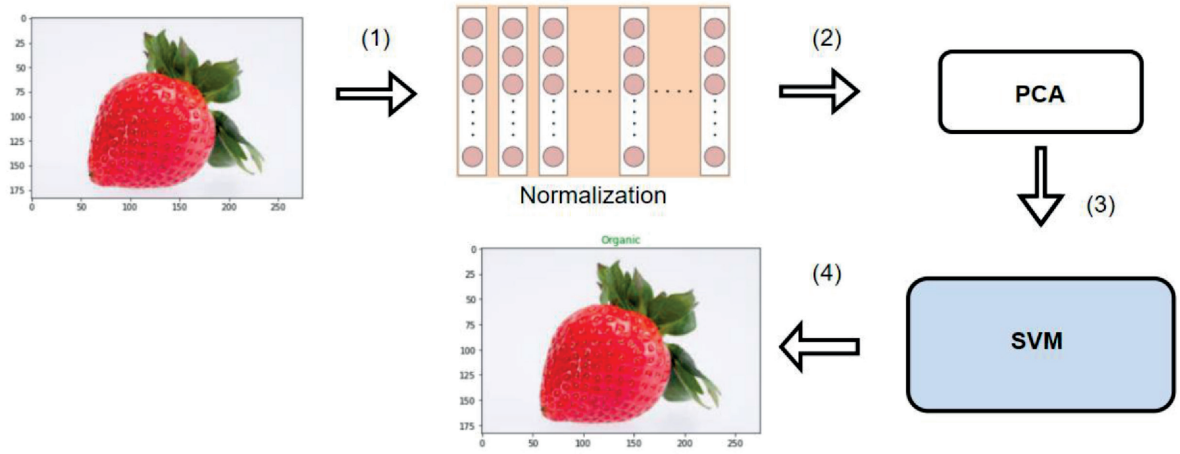

Fig. 3. The overall process of the waste classification method

Scientific Journal of Astana IT University
ISSN (P): 2707-9031
ISSN (E): 2707-904X

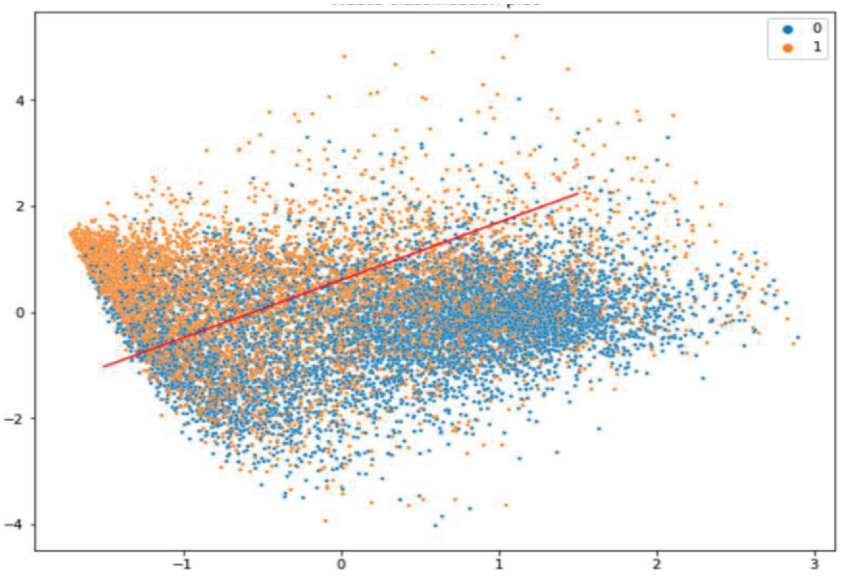

Fig. 4. Waste classification plot

Table 2. Accuracy of each dataset

Dataset	Accuracy
Training set	0.95
Test set	0.88
Cross validation set	0.86

Conclusion

This research paper discussed garbage disposal issues and applied machine learning algorithms to classify waste. To minimize the impact caused by improper waste disposal, we offer an automated system aimed at the correct separation of waste into recycling categories. Was considered two categories of waste: organic and recyclable. The results showed that the classification by the support vector of machines method is an effective approach to solving this problem, reaching 88% accuracy. The dataset is composed of organic and recyclable features. This allows us to identify and classify massive waste disposal data and improve the ecology of the world. In the future, this automated system may be useful for solving problems with waste sorting. Many government agencies are struggling with the consequences of pollution, and this may be the most optimal solution to this problem in the coming centuries. An automated waste distribution system can significantly reduce resource costs. In this work, it is clearly shown how machine learning in the shortest possible time can sort waste automatically in real-time. This work was classified by a dataset of more than 10,000 images, according to the results of SVM, it was found that 56% organic and 44% recyclable waste in our dataset. The result of this work can serve for the further development and automation of the issue of disposal and proper sorting without introducing risk to human health. This SVM model can serve as a solution to the global pollution problem, as well as help with the automation of garbage sorting. Moreover, compared to other research works in the field of machine learning, here we used support vector machines algorithm which also gives us good accuracy.

DOI: 10.37943/AITU.2021.52.74.005

References

1. Shi, C., Tan, C., Wang, T., & Wang, L. (2021). A waste classification method based on a multilayer hybrid convolution neural network. *Applied Sciences*, *11*(18), 8572.
2. Kennedy, T. (2018). OscarNet: Using transfer learning to classify disposable waste. *CS230 Report: Deep Learning. Stanford University, CA, Winter*.
3. Donovan, J. (2016). Auto-trash sorts garbage automatically at the techcrunch disrupt hackathon. *Techcrunch Disrupt Hackaton, San Francisco, CA, USA, Tech. Rep. Disrupt SF*, *2016*.
4. Batinić, B., Vukmirović, S., Vujić, G., Stanisavljević, N., Ubavin, D., & Vukmirović, G. (2011). Using ANN model to determine future waste characteristics in order to achieve specific waste management targets-case study of Serbia.
5. Dong-e, Z., Rui, W., Bao-guo, Z., & Yuan-yuan, C. (2019). Research on Garbage Classification and Recognition Based on Hyperspectral Imaging Technology. *Spectroscopy and Spectral Analysis*, *39*(3), 917-922.
6. Glouche, Y., & Couderc, P. (2013, June). A smart waste management with self-describing objects. In *the Second International Conference on Smart Systems, Devices and Technologies (SMART'13)*.
7. Kennedy, T. (2018). OscarNet: Using transfer learning to classify disposable waste. *CS230 Report: Deep Learning. Stanford University, CA, Winter*.
8. Prashant, B. (2019). SVM Classifier Tutorial. https://www.kaggle.com/prashant111/svm-classifier-tutorial
9. Khanna, D., Sahu, R., Baths, V., & Deshpande, B. (2015). Comparative study of classification techniques (SVM, logistic regression and neural networks) to predict the prevalence of heart disease. *International Journal of Machine Learning and Computing*, *5*(5), 414.
10. Auria, L., & Moro, R. A. (2008). Support vector machines (SVM) as a technique for solvency analysis.