Methods of Estimating Waste Accumulation Rate in Rural Areas of the Lubelskie Voivodship

Tomasz Szul, Krzysztof Nęcka
University of Agriculture in Krakow, Poland

Abstract
The study compared the effectiveness of the rough set theory and artificial neural networks with respect to predicting the rate of waste mass accumulation for recipients in the areas of rural municipalities. Simulations were performed for two variants of input variables. The first of them used all economic, infrastructure and economic indicators as independent variables. The second case was limited only to those whose correlation with the class label attribute exceeded 0.2 and they included: population density, percentage of buildings in the municipality covered by the collection system, the rate of income, and agricultural area. The analysis showed that rough sets’ models generate comparable-quality forecasts of mass waste accumulation rate for rural municipalities, such as artificial neural networks. The developed models are characterized by a high forecast error of about 20%–27%. Further research is needed towards finding effective methods or other conditional attributes that describe the rate of mass accumulation of waste in the areas of rural municipalities.

Keywords: household, waste, forecasting, artificial neural networks, rough sets

Introduction
Amendments to the Act on maintaining cleanliness and order in municipalities revolutionized their waste management system. Municipalities have become owners of waste and they have, therefore, taken over full control of waste management in their territory. Development of a waste management system, in addition to economic criteria, must also take into account social acceptability criteria and eco-efficiency as well as control over the amount of waste generated in a given area. The basis for rational planning of waste management is the so-called unit waste accumulation rate, the correct choice of which is the most important task of the stage of planning the logistics of collecting, sorting and storing waste (Tałałaj 2011). The groups of conditions affecting the amount of waste generated include: economic, social and infrastructural factors. The mere indication of the groups of elements which impact the amount of waste produced is insufficient, because the power of their mutual interaction is unknown (Beigl, Lebersorger, and Salhofer 2008; Beigl et al. 2005). In addition, there are differences in the amount of waste generated with respect to both urban and rural areas. According to data from the Central Statistical Office for 2013, the rate of total mass waste accumulation per capita amounted to 273 kg per person per year in the case of urban areas and it amounted to 121 kg per person per year in the case of rural areas. The rate of waste generated from households amounted to 199 kg per person per year in a city and 97 kg per person per year. Differences between urban and rural areas can also be seen when analyzing the rates’ coefficients of variation, the coefficient of variation is 12% for cities and 33% for rural areas — indicating a large diversity of recipients. Thus the choice of a method enabling development of a model for forecasting the amount of waste generated within a municipality or in households, constituting the basis for planning the economy in the given area, should take into account a number of features.

1. [In the journal European practice of number notation is followed — for example, 36 333.33 (European style) = 36 333.33 (Canadian style) = 36,333.33 (US and British style). — Ed.]
which are predicted to significantly affect the final result. Studies on municipal waste management include the problem of the manner of estimating the amount of waste collected. Many authors have conducted research on the power and direction of the impact of various economic and social factors on the mass of produced or collected waste. Models describing the relationship between socio-economic factors and the properties of waste are frequently based on information from municipalities or Statistical Offices (Bach et al. 2004; Beigl, Lebersorger, and Salhofer 2008; Hage and Soderholm 2008; Hockett, Lober, and Pilgrim 1995; Lebersorger and Beigl 2011; Malinowski et al. 2009; Miller et al. 2009; Passarini et al. 2011; Purcell and Magette 2009). Most of these models were developed for waste management planning at the national level or in urban areas. Models based on the studies are usually multi-factorial. No access to relevant statistics makes it impossible to verify the models in other areas. Analyses at the national (Daskalopoulos, Badr, and Probert 1998; Mazzanti and Zoboli 2008) or regional (Chung 2010) level show an overall relationship between the amount of waste and population density or gross domestic product. These models cannot be applied to waste management planning at lower regional levels (in municipalities). The approaches to estimating the rate of mass accumulation of waste presented in the literature review mainly use statistical methods in the form of linear regression models (den Boer et al. 2010; Malinowski et al. 2009; Sircar, Ewert, and Bohn 2003), the multiple regression (Bach et al. 2004; Malinowski 2013; Nęcka and Szul 2016; Tałałaj 2011), artificial neural networks (Chmielińska-Bernacka and Sidielko 2013) as well as hybrid models that are a combination of these methods (Nęcka and Szul 2014).

Regression analyzes are usually time-consuming and rarely accurately estimate the rate in small areas—especially in rural areas due to lack of data, inaccuracy and uncertainty (den Boer, den Boer, and Jager 2005; Malinowski 2013; Nęcka and Szul 2016; Tałałaj 2011). Therefore, the authors propose a procedure for estimating the rate of mass accumulation of mixed municipal waste in households using the rough set theory (RST) and alternative regression methods using artificial neural networks (ANN) which enable analysis of imprecise, vague and uncertain data.

1 Methodology of research

The paper presents a comparative analysis of the efficiency of the use of the rough set theory and alternative regression in determining the rate of total mass municipal waste ($d_{a1}$) and household waste ($d_{a2}$) accumulation. The data for the implementation of the research was obtained from the Local Data Bank and related to rural municipalities and the rural areas of the urban-rural province of the Lubelskie Voivodship in 2013. The area covered 169 rural and 21 urban-rural municipalities. The value of the rate of total municipal waste accumulation per capita in the rural areas of the region is twice lower than the national average and amounts to 59 kg per person per year and 48 kg per person per year in the case of households with a coefficient of variation for these indicators of 46%. The objects (municipalities) have been described using the following indicators: population density, the average number of people living in a residential building, the percentage of buildings in the municipality covered by the collection system, the administration type of a building, the functional type of a municipality, the rate of income, arable agricultural area, the share of arable land in the use structure, the rate of total mass waste accumulation and the rate of household waste mass accumulation. The study group included municipalities which generated just 10–12 kg of waste per person during a year which was considered not a viable amount, therefore, the objects for which the rate of mass waste accumulation was lower than 30 kg per person were eliminated. Further analyses used indicators for 117 municipalities. In order to verify the admissibility and relevance of the developed models the study group was randomly divided into two sets. The learning set was created with 70% of the observations, while the remainder was a test set. Prognostic models were developed in two variants. In the first one, as input variables were used all of the indicators describing the studied municipalities. The second variant used only four selected attributes whose correlation with the rate of mass waste accumulation was greater than 0.2.
The first method of estimating the rate of mass waste accumulation was a method based on the rough set theory (RST), used to test imprecision, vagueness and uncertainty in the process of analyzing data (Pawlak 1997). Given the facts that the methodology of inference using the rough set theory only takes into account the qualitative nature of the characteristics of objects and the attributes that describe the size of the waste stream are described in quantitative terms—e.g., the density of population, the rate of a municipality income per capita or the given feature share in the whole, integration of rough sets with fuzzy sets in the form of application of evaluated tolerance relation (ETR) was used (Stefanowski and Tsoukas 2001). This enabled the introduction of greater flexibility in data exploration into the rough set theory and a possibility to analyze observations in their quantitative form. The classical RST assumption is based on the concept of the indistinguishability relation as the exact equivalence relation, that is, objects will only be indistinguishable if they have similar attributes (a 0–1 system). The introduction of evaluated tolerance relation into the RST enables the determination of the upper and lower approximation of a set with varying degrees of the indistinguishability relation (d’Amato 2008). This enables the comparison of two sets of data and getting a score between 0 and 1, being the level of the indistinguishability relation. The said interval is a function of belonging derived from the assumptions of the theory of fuzzy sets. The closer the result is to one the more similar (indistinguishable) the objects are in terms of the analyzed attribute, and the closer it is to 0, the more they are distinguishable (Renigier-Bilozor 2008a, 2008b; Renigier-Bilozor and Bilozor 2013). Calculations of the rate of mass waste accumulation were carried out in accordance with the methodology in the paper (Renigier-Bilozor 2008b). Alternative regression models were developed in Statistica 10.0. Selection of the optimum neural network architecture as well as the size of the weights were done using an automatic network designer. The development of the network analyzed neural networks having the following activation functions in the hidden and output layers: identity, logistic sigmoid, hyperbolic tangent and exponential functions. The maximum number of hidden layers was 20. The quality of the developed models was evaluated on the basis of the MAPE error value determined for each set.

2 Findings

The paper is limited to rural municipalities and rural areas of urban and rural municipalities, as previous research shows that the areas are concerned with the estimation of the rate of mass waste accumulation burdened with the greatest error (Nęcka and Szul 2016). The research results presented in the paper have been obtained on the basis of statistical data on the Lubelskie Voivodship, as obtained from the Local Data Bank3 and related to 2013. In the studied year, the rural areas of the voivodship were inhabited by 1158 thousand people. They generated 68,6 thousand tons of waste which accounted for 22,6% of the waste stream in the whole region.

It was decided for the calculations to use the data available in the statistical statements describing the studied municipalities in the rural areas. Two variants of input variables were assumed. In the first one, the following conditional attributes were input: population density, the average number of people living in a residential building, the percentage of buildings in the municipality covered by the collection system, the rate of income, arable agricultural area, the share of arable land in the structure of use and the functional type of municipalities (Bański 2009). In the case of the second variant, only the attributes with the power of correlation with the rate of mass waste accumulation exceeding 0,2 were selected and they included: population density, the percentage of buildings in the municipality covered by the collection system, the rate of income, and arable agricultural area. Table 1 summarizes the characteristics of the variability of parameters characterizing individual objects for which conditional attributes are determined sequentially with symbols from $c_1$ to $c_6$ and class label ones with symbols $da_1$ and $da_2$.

Rates of mass waste accumulation were estimated using the rough set theory (TYPE) in the application with evaluated tolerance relation (ETR). An information system (a class label table) consisting of 82 objects (the learning set) was developed. The objects included in the learning set

are presented in the form of a class label table in which the characteristics of a given municipality are conditional attributes (respectively, from variant I and variant II) and the rates of total mass waste accumulation and household waste accumulation are class label attributes $da_1$, $da_2$.

Four information systems were developed:

- conditional attributes from the first variant — class label attribute $da_1$
- conditional attributes from the second variant — class label attribute $da_1$
- conditional attributes from the first variant — class label attribute $da_1$
- conditional attributes from the second variant — class label attribute $da_1$

Calculations carried out in accordance with the methodology (Renigier-Biłozor 2008b) resulted in a core comprising 76 representative class label rules. After the emergence of representative class label rules, it was possible to determine the rate of mass waste accumulation. For this purpose, the municipalities from the test set were used. In the case of the set of 35 municipalities constituting the test objects, conditional attributes were assumed for the first and second variant, then applying the evaluated tolerance relation (ETR) to determine the degrees of belonging to class label rules and to verify which of the analyzed municipalities belonged to which of the selected class label rules to the greatest degree. The final step of calculations was to calculate the relative error of estimation of the various variants and class label attributes, as summarized in table 3. The average values of relative estimation errors of the rates of the total waste and household accumulations are similar, ranging from 21 (the first variant) to 30% (the second variant).

The artificial neural networks (ANN) available in Statistica 10.0 were used for the development of alternative regression models. As this program makes it possible to develop networks with various configurations, it was decided to automatically select the optimal network configuration. The developed networks used both the Multi-layer Perceptron (MLP) and radial networks. In the networks developed, an automatic designer had specific conditional and class label attributes. A limitation was also assumed for the minimum number of hidden layers to be 2 and the maximum to be 20. During the network learning, the usefulness of the following learning algorithms was verified: the steepest descent, conjugate gradients and the variable Broyden-Fletcher-Goldfarb-Shanno metrics (BFGs). The program also verified how the type of neuron activation function of hidden layers and the output layer will affect the quality of the model. During the tests, the usefulness of following the activation functions was verified: the identity, logistic sigmoid, hyperbolic tangent and exponential. For individual variants of combinations of input variables, 200 networks were developed and 5 of them were selected for further analysis, those characterized by the smallest forecast errors. The indicators characterizing the quality of the developed ANN models are summarized in table 2.

The conducted analyses prove that, using an ANN, one can obtain forecasts of the rate of mass waste accumulation with an average relative error of 20%–27%. It was observed that both the rate of total waste accumulation and the household waste accumulation are affected by errors at a comparable level. The MAPE error values for the learning set were on average lower by more than 3% than for the test set. In the case of all of the trials, the most effective were the networks developed on the basis of the multi-layer perceptron (MLP). No clear answer considering the optimal network architecture was obtained since the results obtained for the network of only 4 and 19 neurons in

<table>
<thead>
<tr>
<th>Type of municipality</th>
<th>Measure</th>
<th>$c_1$</th>
<th>$c_2$</th>
<th>$c_3$</th>
<th>$c_4$</th>
<th>$c_5$</th>
<th>$c_6$</th>
<th>$da_1$</th>
<th>$da_2$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Urban-rural</td>
<td>Mean</td>
<td>94.2</td>
<td>4.4</td>
<td>65.9</td>
<td>341.0</td>
<td>4.3</td>
<td>82.2</td>
<td>105.5</td>
<td>79.1</td>
</tr>
<tr>
<td></td>
<td>Variability (%)</td>
<td>66.2</td>
<td>44.5</td>
<td>24.6</td>
<td>36.8</td>
<td>32.0</td>
<td>11.7</td>
<td>46.5</td>
<td>44.0</td>
</tr>
<tr>
<td>Rural</td>
<td>Mean</td>
<td>51.4</td>
<td>3.4</td>
<td>58.8</td>
<td>241.0</td>
<td>6.5</td>
<td>85.9</td>
<td>56.1</td>
<td>45.9</td>
</tr>
<tr>
<td></td>
<td>Variability (%)</td>
<td>47.0</td>
<td>12.8</td>
<td>36.7</td>
<td>38.1</td>
<td>43.9</td>
<td>7.9</td>
<td>55.2</td>
<td>53.7</td>
</tr>
</tbody>
</table>

Note: $c_1$ — population density (persons/km²); $c_2$ — mean number of persons living in a residential building (persons/building); $c_3$ — percentage of buildings in the municipality covered by waste collection scheme; $c_4$ — income indicator (own revenues of municipalities — shares in the taxes constituting the revenues of the state budget, revenue tax from natural persons) (PLN per person per year); $c_5$ — area of arable land (hectares); $c_6$ — proportion of arable land in the structure of land use (%); $da_1$ — overall mass waste accumulation indicator (kg per person per year); $da_2$ — mass waste accumulation indicator from households (kg per person per year).
the hidden layers were comparable. Selecting only the three most strongly correlated variables as input variables resulted in a slight improvement in the quality of forecasts.

Table 3 summarizes the indicators characterizing the forecasting models developed using the rough set method and artificial neural networks.

The lowest value of the average relative forecast error for the rate of mass waste accumulation for all recipients in the rural areas of the Lubelskie Voivodship designated for the test set were obtained for the model basing on the rough set theory, and it was at the level of 21.2%. This model used all seven conditional attributes as input variables. The use of the same set of input variables in the ANN model resulted in an almost 3% increase of the forecast error and at the same time a significant increase in volatility.

In order to illustrate the course of the forecast error for each model using rough sets and artificial neural networks AFE distribution functions were developed, as shown in figure 1. As is apparent from the figures, the lowest share of AFEs with large values is generated by the rough sets’ method with the first variant of the input variables used. Forecasting both the total value of the mass waste accumulation rates and its level for households in rural municipalities, the maximum error values do not exceed 60%. The ANN models developed to predict the total value of the waste accumulation rate regardless of the selected variant of input variables generated very similar distribution of errors. Only the use of the second set of input variables in the rough set theory resulted in increasing the forecast error and its maximum values reached almost 100%. Modeling the value of da2 for households proved some advantage of the ANN model (model 4) over models 2 and 3 but model 1 still generated the lowest number of major APE values but at the same time it also characterized the lowest number of forecasts of the greatest accuracy.
Conclusions

The analyses carried out showed that rough sets models generate a bit more precise forecasts of the mass waste accumulation rate for rural municipalities than artificial neural networks. Instead they are very sensitive to a set of input variables which was not observed in the case of ANNs. It was observed that higher-quality predictions of mass waste accumulation rate for rural recipients are recorded in the case of networks developed on the basis of the Multi-layer Perceptron MLP than radial networks. No clear answer concerning the optimal network architecture was obtained since comparable results were obtained for networks of 4 and 19 neurons in hidden layers. The models developed are characterized by a high forecast error of about 20%–27% and it can be considered that they are comparable methods used to predict mass waste accumulation. Further studies will be aimed at finding better conditional attributes to be easily collected by local governments.

References


den Boer, E., J. den Boer, and J. Jäger. eds. 2005. Planowanie i optymalizacja gospodarki odpadami. Podręcznik prognozowania ilości i jakości odpadów komunalnych oraz oce-
ny zgodności systemów gospodarki odpadami z zasadami zrównoważonego rozwoju. Wrocław – Kamieniec Wrocławski: Oddział Dolnośląski PZiTS; WAMECO.


