Rough Set-based Model of Waste Management Systems towards Circular City Economies

Ivan Henderson V. Gue[a,*], Neil Stephen Lopez[a], Anthony S. F. Chiu[b], Aristotle T. Ubando[a], Raymond R. Tan[c]

[a] Mechanical Engineering Department, De La Salle University, 2401 Taft Avenue, Malate, Manila, Philippines
[b] Industrial Engineering Department, De La Salle University, 2401 Taft Avenue, Malate, Manila, Philippines
[c] Chemical Engineering Department, De La Salle University, 2401 Taft Avenue, Malate, Manila, Philippines
 ivan.gue@dlsu.edu.ph

Cities play a significant role on the fate of materials in economy-wide flows. Their sustainability is integral to the future of resource utilization. Circular city economies are recommended for the future of sustainable cities. Shifting towards a circular economy is challenging as urban metabolism is intertwined with city characteristics. The effectiveness of action plans is vital towards the shift through insight derivation of past city-level data. Rough set-based model can draw insights from city-level data in interpretable form, facilitating communication between analysts and city planners. This approach generates if-then rules based on the city-level data. This work generated 14 if-then rules from data of 100 cities using rough set theory. The model identified 5 relevant city characteristics from the 13 characteristics available from the data. The relevant characteristics are demographic, education, ease of doing business, income inequality, and tourism. The metric on waste management for cities was selected as the decision attribute of the model as it is the most relevant to circularity. The model attained a classification accuracy of 94 %. Specific if-then rules achieved coverage as high as 64 %, allowing ease of analysis. The rules suggest the level of development as the delineation between waste management system performances. The findings highlight the relevance of future studies on circular economy of developing countries. Results of this work provide insights on critical features that occur in more sustainable cities, which can be used to plan future circular city economies.

1. Introduction

Cities encompass more than half of the global population (United Nations, 2018) and constitute 80 % of the global GDP (World Bank, 2020). The population level and economic performance prompts high material utilization. Material consumption of cities accounts for more than half of the global sum (IRP, 2018), highlighting the influence of cities to future sustainability goals. Sustainable cities are integral towards resource decoupling. Adoption of circular economy (CE) among cities is recommended for their future sustainability (IRP,2018). CE's philosophy is applicable among cities and other scales of economic systems such as companies and industrial parks (Kirchherr et al., 2017). Circular systems are argued to exhibit carbon footprint and material footprint reductions (Geng et al., 2019).

Cities are complex economic systems, exhibiting multifaceted characteristics. Such as in the work of Nakamura (2019), analysis of land price included analysis on entrepreneurial, environment, and social factors. The complexities resulted to the concept of urban metabolism where technical, social, and economic factors affect material flows among cities (Wolman, 1965). Understanding the interplay of factors and urban metabolism is significant for city planning towards sustainability (Kennedy et al., 2011). Critical assessment on the prerequisites of sustainable cities is integral to action plan development. Elliot et al. (2019) recommends analysing the cause-and-effect of urban metabolism for sound decision making. Remøy et al. (2019) added that evaluating the cause-and-effect is integral for the transition towards CE.

Complexities of urban metabolism yields difficulties in its evaluation and estimation. Machine learning (ML) provides a resolution to the challenges of estimation. Its prominence has been recognized for sustainable development applications (Gue et al., 2020). Nosratabadi et al. (2020) noted that there is an exponential growth

in publication on the applications of ML on city development. Feng and Xu (1999) applied a hybrid Artificial Neural Network (ANN) for decision making in city planning. Reades et al. (2019) used ML to identify gentrification. Khoshnava et al. (2020) applied ANN in estimating the relationship between green infrastructures and the economy.

The prevailing limitation of ML in urban planning is its 'black box' approach (Wagner and de Vries, 2019). The 'black box' limitation impedes acceptance of the models among stakeholders. Rudin (2019) recommends interpretable models as the solution to the 'black box' limitation. Interpretability is critical for high-stakes decision making (Carvalho et al., 2019). Doshi-Velez and Kim (2017) argued the need of interpretability for cases of significant consequences and cases of difficult real-world validation. City planning is a high-stakes decision making as it affects a significant portion of the population. City planning is difficult to validate because of the scale and multifaceted nature of cities. Interpretable models are essential in city planning. Gue et al. (2021) used an interpretable model in the depiction of causal network maps for urban circular economies.

Rule induction techniques are interpretable ML models that generate if-then rules in modelling datasets (Grzymala-Busse, 2009). The rules are easily interpretable as causal relationships enable effective human understanding (Holzinger et al., 2019). Decision Trees (DT) is a prominent rule induction technique, yielding a tree-like structure in the prediction model. Saldivar-Sali (2010) used Decision Trees (DT) in classifying cities according to climate, GDP, and population. The tree-like structure of DT predicts through a cascade of if-then rules. The cascaded approach becomes difficult to comprehend as additional branches are included.

Rough Set Theory (RST) is an alternative rule induction technique that are easier to comprehend, using simple if-then rules. RST generates rules according to imprecise boundaries (Pawlak, 1982). The technique has the advantage of capturing rules for entirety of the data. Stakeholders may omit rules that does not meet a specified accuracy or coverage. RST is a recognized technique for decision support systems. Aviso et al. (2008) applied the technique as a in improving the industry's environmental performance. Literature has also applied RST in the evaluation of cities and countries. Szul et al. (2017) estimated household waste generation among rural areas. Their model generated 40 decision rules. He et al. (2018) evaluated the clean energy development among countries. Lopez et al. (2021) identified socio-economic traits affecting carbon footprint of countries.

As CE is a conglomeration of existing ideologies, recognizing historical patterns derives insights for city planning. City planning for CE requires holistic analysis of the cities' complexities. The complexities entail the use of ML models to derive patterns from historical data to guide future action plans. The level of interpretability is important for decision making. RST is a rule induction technique that is interpretable with simple if-then rules. Literature has yet formulated an RST of if-then rules on the circular performance of cities. This work generates a rough set-based model on city characteristics affecting their circular performance. Metric on waste management is selected as the indicator of circular performance. The significance of this work is the visualization of if-then rules on the critical historical patterns of sustainable cities. Insights drawn from the findings can support the development of key action plans for circular city economies The generated rules also enable ease of comprehension when relaying historical patterns to key decision makers.

This section provided an overview of the problem and the study's objective. Section 2 details the methodology which includes a brief discussion of RST and a description of the city-level data. Section 3 shows and discusses the findings. Section 4 provides the conclusion of this work.

2. Methodology

Rules for modelling city characteristics were formulated using RST. This work utilized characteristics of 100 cities. The dataset was divided into a 70:30 ratio for training and testing. The dataset was trained and validated through the ROSETTA software developed by Øhrn (1999).

RST is a pattern recognition technique utilizing attribute discernibility for rule formation (Pawlak, 1982). Hvidsten (2013) provides a step-by-step discussion on the technique's computational procedure. Continuous variables of the dataset are initially discretized, resulting to a binary classification for this work. The discernibility between conditional attributes and the decision attribute is assessed. Reducts are then formed from the discernibility. Reducts are sets of conditional attributes that are sufficient to describe the decision attribute. RST then generates if-then rules from a selected reduct.

Data on city characteristics are obtained from the Sustainable Cities Index (SCI) of Arcadis (2016), representing 100 cities. As of writing, the numerical data of the SCI dataset became inaccessible. The case study therefore presents a demonstrative procedure of rough set-based modelling. The SCI dataset comprise of three main city characteristics which are people, profit, and planet. The three characteristics has its own sub-indicators, tallying to 20 sub-indicators. This study utilized the 13 sub-indicators of people and profit as the conditional attribute. This study considered the 7 sub-indicators of planet as candidates for the decision attribute. The sub-indicator 'waste management' significantly represents the critical component of CE (Kŭdela et al., 2020). 'Waste

management' was then chosen as the decision attribute. The other sub-indicators under the planet characteristic were not selected as their relevance to CE is minimal.

Scoring by SCI is through normalization of quantifiable characteristics. The normalization process is based on the highest and lowest score of the 100 cities. The scores were benchmarked relative to the top and lowest performance levels. Sub-indicators closer to 1 indicate that the performance level is closer to the top benchmark while the opposite is true for 0. The relative basis is used to discretize the numerical representations to binary classifications. Scores between 0.5 to 1 are classified as '1'. This category indicates that the city's attribute is closer to the top benchmark than the lowest. Scores between 0 to 0.5 are classified as '2'. This category indicates the city's attribute is closer to the lowest than the top benchmark. The discretization therefore follows benchmarking as a performance indicator. Insights drawn are bounded on how well the cities perform relative to the concurrent extremes. After discretization, RST identified the dataset's reducts. The ROSETTA software searches reducts through a Genetic Algorithm (GA) approach. The software allows a threshold to be set where reducts supporting a threshold percentage of the dataset are also determined. The preferred reduct was then used for rule generation.

3. Results and Discussion

The GA component identified multiple reducts. The reducts are sets of conditional attributes that are adequate to describe the decision attribute. The reduct with the least number of attributes was selected for ease of analysis. Table 1 details the description of each attribute and the instances of their classifications.

Table 1: Selected attributes with their description and instances

Notation	Metric	Description	Instances
DE	Demographics	Described by the dependency ratio. Classification of '1' indicates larger population ratio capable to join the labour force.	'1' – 40 '2' – 60
ED	Education	Described by the level of education. Classification of '1' indicates better educational system.	'1' – 59 '2' – 41
II	Income Inequality	Described by the Gini coefficient. Classification of '1' indicates equally distributed income.	'1' – 58 '2' – 42
EB	Ease of Doing Business	Described by the Ease of Doing Business Index. Classification of '1' indicates better business environment.	'1' – 68 '2' – 32
TO	Tourism	Described by the number of tourists per year. Classification of '1' indicates higher tourist rate.	'1' – 23 '2' – 77
WM	Waste Management	Described by the amount of landfill, recycling, and wastewater treated. Classification of '1' indicates better waste management	'1' – 59 '2' – 41

Figure 1 depicts the confusion matrix. The model did not capture all the objects in the dataset as the conditional attributes of one object cannot be looked up. This limitation resulted to a no classification instance, denoted as '*'. Overall, the rules have correctly classified 94 % of the dataset, indicating satisfactory classification performance.

		Predicted				Recall	FN Rate
		1	2	*			
Actual	1	56	2	1		94.92%	5.08%
	2	3	38	0		92.68%	7.32%

*No classification

					Overall Accuracy
Precision	94.92%	95.00%			
FP Rate	5.08%	5.00%			94.00%

Figure 1: Confusion matrix of the generated rules from the SCI dataset

The reduct generated 22 rules which were then conglomerated into 14 if-then rules. Table 2 enumerates the 14 if-then rules. The attributes DE, ED, II, EB, and TO are the conditional attributes and represent the LHS of the if-then rules. The attribute WM is the decision attribute and represent the RHS of the if-then rules. For example, Rule 2 states if DE is 'DNC', ED is '2', II is '2', EB is '2', TO is '2', then WM is '1' or '2'.

TO has the highest instance of 'DNC' among the five attributes. The instances are caused by the attribute's skewed distribution, as indicated in Table 1. Rules 2, 7, and 11 have two decision classes for WM. Their prediction of the decision attribute is either '1' or '2'. Selection of which decision attribute is dependent on preference. This work selected the decision attribute with the higher RHS accuracy. For example, the RHS accuracy of Rule 2 for '2' is 83 %. Objects of the corresponding conditional attributes are predicted to have a WM of '2'.

The LHS support indicates the objects captured by the rule while the LHS coverage is its percent share. Rule 1 has the highest LHS support constituting 38 of the 100 objects. The RHS support indicates the number of objects with the corresponding decision class. The RHS coverage is the percent share with reference to the decision class' total instance. Rule 1 has the highest RHS coverage of '1', constituting 64 % of the decision class. Rule 2, on the other hand, has the highest RHS coverage of '2', encompassing 24 %.

Table 2: Generated rules from the SCI 2016 dataset

Rule	DE	ED	II	EB	TO	WM	LHS Support	RHS Support	RHS Accuracy	LHS Coverage	RHS Coverage
1	2	1	DNC	1	DNC	1	38	38	1.00	0.38	0.64
2	DNC	2	2	2	2	1, 2	12	2, 10	0.17, 0.83	0.12	0.03, 0.24
3	1	1	2	1	DNC	1	9	9	1.00	0.09	0.15
4	2	2	1	2	2	2	8	8	1.00	0.08	0.20
5	1	2	1	DNC	DNC	2	6	6	1.00	0.06	0.15
6	1	1	1	1	2	1	6	6	1.00	0.06	0.10
7	2	2	1	1	DNC	1, 2	5	3, 2	0.60, 0.40	0.05	0.05, 0.05
8	1	2	2	2	1	2	4	4	1.00	0.04	0.10
9	1	1	2	2	2	2	3	3	1.00	0.03	0.07
10	1	2	2	1	DNC	2	3	3	1.00	0.03	0.07
11	1	1	1	1	1	1, 2	2	1, 1	0.50, 0.50	0.02	0.02, 0.02
12	2	2	2	1	2	2	2	2	1.00	0.02	0.05
13	2	1	2	2	2	2	1	1	1.00	0.01	0.02
14	2	2	2	2	1	2	1	1	1.00	0.01	0.02

*'DNC' indicates attribute's value is irrelevant

Classification of '1' are countries with relatively better waste management systems. Rule 1 describes 64 % of such instances. The conditional attribute for the rule is attaining a '1' classification for ED and EB, reflecting good environment for education and business operations. This is coherent with the findings of Smejkalová et al. (2020) where education and economy are significant to waste management. Complimentary to the two attributes is the '2' classification of DE, indicating low score for demographics. The low score of DE is reflective of cities from developed countries as they exhibit lower dependency ratio (Engelgau et al., 2011). Rules 3 and 6 also describe WM of '1'. The two rules are also indicative of '1' classification for ED and EB. The aggregated RHS coverage of Rules 1, 3, and 6 is 90 %. Classification of the three rules' conditional attribute are reflective of cities from developed countries.

Classification of '2' are countries with relatively poor waste management system. Rules 2, 4, and 5 describe 59 % of such instances. The three rules have the conditional attribute ED set as '2', indicating the education system on the lower half. The attributes' '2' classification is also reflected on the discussion of Diaz (2017) where absence of education impedes participation of proper waste management. EB is also set to '2' except for Rule 5. The score of both attributes reflects characteristics of cities from developing countries.

Table 3 shows a sample illustration of the RST's prediction. The table demonstrates the classification of the six cities according to the 14 if-then rules. The findings indicate that the difference of waste management system is delineated by the difference between developed and developing countries. The finding is coherent with the review of Halog and Anieke (2021) as developing countries have notable mismanagement of waste. Waste management is a component of CE. The decision attribute can be a depiction on the future success rate of circular city economies. The rules generated by RST suggest an impending challenge of CE among cities of developing countries. As highlighted in the review of Halog and Anieke (2021), CE is a perspective scarcely prioritized among developing countries.

Table 3: Sample demonstration of the RST model

City	Activated Rule	Conditional Attribute					Decision Attribute
		DE	ED	II	EB	TO	WM
Amsterdam	1	2	1	1	1	1	1
Antwerp	1	2	1	1	1	2	1
Baltimore	3	1	1	2	1	2	1
Manila	2	2	2	2	2	2	2
Sau Paulo	2	1	2	2	2	2	2
Kolkata	4	2	2	1	2	2	2

4. Conclusions

This work generated if-then rules of city data characterizing proper waste management using machine learning approach. The rules attained a classification accuracy of 94 %. Rule 1 covers 64 % of better performing cities while Rules 2, 4, and 5 captures 59 % of poor performing cities. Findings of this work highlight the relevance of studies on CE for cities from developing countries. The rules show that the delineation between waste management system performance is the difference between level of development. The rules for poorer waste management indicate poor education and business environment. Transition towards circular city economies will need to consider challenges of poor education and business environment.

The dataset encompassed all cities regardless of regional differences. Regional differences, such as continents, portray distinction in urban metabolism. Future works may consider certain regional distinctions for generation of distinct rules. City types, such as population dense cities, may exhibit unique rules as well. Future work may consider standard city types as distinction for rule generation. The classification of this work was based on benchmarking relative to concurrent performance levels. The ranges were not based on a standard classification. Future works may consider investigating appropriate classification ranges. The case study demonstrated a rough set-based modelling for circular city economies. Future works may consider other city-level datasets, following the demonstrated procedure.

References

Arcadis, 2016, Sustainable cities index <https://images.arcadis.com/media/0/6/6/%7B06687980-3179-47AD-89FD-F6AFA76EBB73%7DSustainable%20Cities%20Index%202016%20Global%20Web.pdf> accessed 23.12.2020.

Aviso K.B., Tan R.R., Culaba A.B., 2008, Application of rough sets for environmental decision support in industry, Clean Technologies and Environmental Policy, 10(1), 53-66.

Carvalho D.V., Pereira E.M., Cardoso J.S., 2019, Machine learning interpretability: a survey on methods and metrics, Electronics, 8(8), 832.

Diaz L.F., 2017, Waste management in developing countries and the circular economy, International Solid Waste Association, 35(1), 1-2.

Doshi-Velez F., Kim B., 2017. Towards a rigorous science of interpretable machine learning, <arxiv.org/abs/1702.08608> accessed 07.06.2021.

Elliot T., Babí Almenar J., Niza S., Proença V., Rugani B., 2019, Pathways to modelling ecosystem services within an urban metabolism framework, Sustainability, 11(10), 2766.

Engelgau M.M, El-Saharty S., Kudesia P., Rajan V., Rosehouse S., Okamoto K., 2011, Capitalizing on the demographic transition: tackling noncommunicable diseases in South Asia, World Bank, Washington, USA.

Feng S., Xu L.D., 1999, Hybrid artificial intelligence approach to urban planning, Expert Systems, 16(4), 248–261.

Geng Y., Sarkis J., Bleischwitz R., 2019, How to globalize the circular economy, Nature, 565(7738), 153–155.

Grzymala-Busse J.W., 2009, Rule Induction, Data Mining and Knowledge Discovery Handbook, 249–265.

Gue I.H.V, Ubando A.T., Tseng M.-L., Tan R.R., 2020. Artificial neural networks for sustainable development: a critical review, Clean Technologies and Environmental Policy, 22(7), 1449–1465.

Gue I.H.V., Tan R.R., Ubando A.T., 2021, Causal network maps of urban circular economies, Clean Technologies and Environmental Policy, DOI: 10.1007/s10098-021-02117-9

Halog A., Anieke S., 2021, A review of circular economy studies in developed countries and its potential adoption in developing countries, Circular Economy and Sustainability, DOI: 10.1007/s43615-021-00017-0

He Y., Pang Y., Zhang Q., Jiao Z., Chen Q., 2018, Comprehensive evaluation of regional clean energy development levels based on principal component analysis and rough set theory, Renewable Energy, 122, 643-653.

Holzinger A., Langs G., Denk H., Zatloukal K., Müller H., 2019, Causability and explainability of artificial intelligence in medicine, WIREs Data Mining and Knowledge Discovery, 9(4), 1–13.

Hvidsten T., 2013, A tutorial guide to the ROSETTA system: a rough set toolkit for analysis of data, <trhvidsten.com/docs/ROSETTATutorial.pdf> accessed 31.05.2021.

IRP, 2018, The weight of cities: resource requirements of future urbanization, <resourcepanel.org/reports/weight-cities> accessed 24.05.2021.

Kennedy C., Pincetl S., Bunje P., 2011, The study of urban metabolism and its applications to urban planning and design, Environmental Pollution, 159(8–9), 1965–1973.

Kirchherr J., Reike D., Hekkert M., 2017, Conceptualizing the circular economy: An analysis of 114 definitions, Resources, Conservation and Recycling, 127(8), 221–232.

Khoshnava S.M., Rostami R., Zin R.M., Kamyab H., Abd Majid M.Z., Yousefpour A., Mardani A., 2020, Green efforts to link the economy and infrastructure strategies in the context of sustainable development, Energy, 193, 116759.

Kůdela J., Šomplák R., Smejkalová V., Nevrlý V., Jirásek P., 2020, The potential for future material recovery of municipal solid waste: inputs for sustainable infrastructure planning, Chemical Engineering Transactions, 81, 1219-1224

Lopez N.S., Mouy M., Africa A.D., 2021, Uncovering the significant socio-economic attributes of low- and high-emission countries using rough sets, Clean Technologies and Environmental Policy, DOI: 10.1007/s10098-021-02067-2

Nakamura H., 2019, Relationship among land price, entrepreneurship, the environment, economics, and social factors in the value assessment of Japanese cities, Journal of Cleaner Production, 217, 144-152.

Nosratabadi S., Mosavi A., Keivani R., Ardabili S., Aram F., 2020, State of the art survey of deep learning and machine learning models for smart cities and urban sustainability, Engineering for Sustainable Future, 228–238.

Øhrn A., 1999, Discernibility and rough sets in medicine: tools and applications, PhD Thesis, Norwegian University of Science and Technology, Trondheim, Norway.

Pawlak Z., 1982, Rough sets, International Journal of Computer and Information Sciences, 11(5), 341–356.

Reades J., De Souza J., Hubbard P., 2019, Understanding urban gentrification through machine learning, Urban Studies, 56(5), 922–942.

Remøy H., Wandl A., Ceric D., Van Timmeren A., 2019, Facilitating circular economy in urban planning, Urban Planning, 4(3), 1–4.

Rudin C., 2019, Stop explaining black box machine learning models for high stakes decisions and use interpretable models instead, Nature Machine Intelligence, 1(5), 206–215.

Saldivar-Sali A., 2010, A global typology of cities: classification tree analysis of urban resource consumption, MS Thesis, Massachusetts Institute of Technology. Massachusetts Institute of Technology, Cambridge, Massachusetts.

Smejkalová V., Šomplák R., Rybová K., Nevrlý V., Rosecký M., Burcin B., Kučera T., 2020, Waste production and treatment modelling for EU member states, Chemical Engineering Transactions, 81, 691-696

Szul T., Knaga J., Nęcka K., 2017, Application of rough set theory to establish the amount of waste in households in rural areas, Ecological Chemistry and Engineering S, 24(2), 311–325.

United Nations, 2018, World urbanization prospects: the 2018 revision, <population.un.org/wup> accessed 18.12.2020.

Wagner M., de Vries W.T., 2019, Comparative review of methods supporting decision-making in urban development and land management, Land, 8(8), 123.

World Bank, 2020, Urban development overview, <worldbank.org/en/topic/urbandevelopment/overview> accessed 18.12.2020.

Wolman A., 1965, The metabolism of cities, Scientific American, 213(3), 178–190.