ARTICLE IN PRESS

Waste Management xxx (2018) xxx–xxx

Contents lists available at ScienceDirect

Waste Management

journal homepage: www.elsevier.com/locate/wasman

Review

Data analytics approach to create waste generation profiles for waste management and collection

Harri Niska [a,*], Ari Serkkola [b]

[a] Department of Environmental and Biological Sciences, University of Eastern Finland, P.O. Box 1627, FI-70211 Kuopio, Finland
[b] Department of Built Environment, Aalto University, P.O. Box 14100, FI-00076 Aalto, Finland

ARTICLE INFO

Article history:
Received 24 October 2017
Revised 21 February 2018
Accepted 23 April 2018
Available online xxxx

Keywords:
Waste monitoring
Waste generation
Data analytics
Data mining
Machine learning
Cluster analysis

ABSTRACT

Extensive monitoring data on waste generation is increasingly collected in order to implement cost-efficient and sustainable waste management operations. In addition, geospatial data from different registries of the society are opening for free usage. Novel data analytics approaches can be built on the top of the data to produce more detailed, and in-time waste generation information for the basis of waste management and collection. In this paper, a data-based approach based on the self-organizing map (SOM) and the k-means algorithm is developed for creating a set of waste generation type profiles. The approach is demonstrated using the extensive container-level waste weighting data collected in the metropolitan area of Helsinki, Finland. The results obtained highlight the potential of advanced data analytic approaches in producing more detailed waste generation information e.g. for the basis of tailored feedback services for waste producers and the planning and optimization of waste collection and recycling.

1. Introduction

From a policy perspective, there are two main objectives in waste management that can be identified, which are (i) the reduction of waste generation and (ii) the reduction of resource use, with the objective to maximize 'recovery, reuse, and recycling' (RRR), in order to achieve sustainable waste management. These objectives are built-in the waste hierarchy, which has become a widely accepted guideline for waste management operations (the directive 2008/98/EC). To meet the aforementioned waste management policy new information based means are needed to support waste management and recycling operations (e.g. source and material separation, collection).

In the field of waste management, the amount and composition of waste comprise the basic information to be monitored. The waste generation data is however monitored and compiled mainly in regional and annual resolution and coverage, which in statistical perspective may result in an incomplete information picture in terms of high aggregation level, data frequency and spatial dimension (Eurostat, 2001) and may thus limit to make decisions concerning single waste producers and geographical areas. Spatially and temporally accurate data is therefore necessary to achieve deeper understanding of waste generation behavior and to implement feasible waste management policies and operation at site and local levels.

The situation appears about to change, however, as extensive container level monitoring data on waste generation are increasingly gathered e.g. through transportation control systems (TCS) and sensor networks. Furthermore, in parallel with the open data strategies and initiatives (e.g. INSPIRE 2007/2/EC and PSI 2003/98/EC) different registries of society (public data) are opening for free usage. Spatially and timely accurate monitoring data, and its fusion with other information, enables the building of new data-driven insights to the planning and operation of waste and material flows. This advances the reduction of wastes and the use of natural resources. Due to extensive data, advanced algorithms based on machine learning techniques are increasingly needed in implementing efficient analytics. Additionally, the size of the data requires the use of distributed computing techniques and platforms for performing the state-of-the art analytics in real-world waste management applications (see e.g. Bilal et al., 2016).

The aim of the paper is to introduce the possibilities of advanced data analytics for creating useful information for waste management. We examine data based approaches in creating waste generation information based on waste monitoring data and other information. A computational approach based on the self-organizing map (SOM) and the k-means clustering algorithm is proposed for creating waste generation type profiles. The approach is demonstrated using the extensive the weighing data

* Corresponding author.
E-mail address: harri.niska@uef.fi (H. Niska).

https://doi.org/10.1016/j.wasman.2018.04.033
0956-053X/

Please cite this article in press as: Niska, H., Serkkola, A. Data analytics approach to create waste generation profiles for waste management and collection. Waste Management (2018), https://doi.org/10.1016/j.wasman.2018.04.033

on municipal solid waste (MSW) collection, which consists of measurements from nearly 200 000 pick-up events of deep collection containers and front load containers in the period January 2013–March 2015, in the metropolitan area of Helsinki, Finland.

The paper is organized as follows. First, the methods of data analytics and their potential applications to waste generation monitoring data are examined. This is followed by the presentation of the case study, including the experimental data and its processing, the computational methods and the achieved results. Finally, the conclusions and recommendations for future work of waste monitoring data based analytics are laid out.

2. Waste monitoring based data analytics

In principle, data analytics aims at examining raw data to uncover hidden patterns, correlations and other insights. Waste management utilities and companies may apply analytics to describe, predict and improve their operation from available monitoring data and other information. For example, data analytics can be used to analyse and predict waste generation in time, in a region or in a specific waste producer group. Waste generation information can be used as the basis of operational and strategic planning of waste management e.g. collection and transportation. Furthermore, analytics can be used as the basis of waste producer analysis, e.g. to identify exceptional waste generation behavior, and to profile and classify waste producers.

At theoretical level, data analytics process can be described e.g. using the Knowledge Discovery in Databases (KDD) process (Fayyad, 1996). The stages of KDD data processing can be divided into data selection, preprocessing, transformation, data mining and interpretation (Fig. 1). In that process, data mining plays a crucial role, containing a set of data analysis methods, which can be applied to different purposes of analyzing data. The most relevant means are descriptive and predictive modelling. Descriptive modelling explores intrinsic properties of the data e.g. using density estimation and cluster analysis, whereas predictive modelling aims to build a model for estimating output variable(s) from a set input variables (e.g. Hand et al., 2001).

The term 'advanced analytics' is typically related to emerging fields related to data mining, such as applying machine learning to learn and predict complex behavior. Machine learning algorithms include artificial neural networks (ANNs) and support vector machines (SVM). SVM models, in particular, have shown to yield good prediction accuracy, beating the conventional ANN models in performance and robustness. Relatively new emerging area in the domain of machine learning is Deep Learning (LeCun et al., 2015), which attempts to create (ANN-based) mathematical models for learning complex representations from large-scale (unlabeled) data. Various deep learning architectures such as Deep Belief Networks (DBN), Long Short-Term Memory (LSTN) and recurrent neural networks (RNNs) are currently studied, and are of interest in case of extensive and complex data.

Compared to the conventional statistics, machine learning algorithms, have capacity to find unknown, complicated functional relationships from extensive data, without prior information about phenomenon. This is a benefit compared to mechanistic (physical) based modelling, but may however lead to too complex, black-box models with limited mechanistic interpretation. The lack of mechanistic understanding may limits the acceptance of the purely data-based models in real-world applications, and therefore simplicity, and interpretation should be among the major criteria in data-based generation of models for applications.

Based on the aforementioned issues, the main steps of data analytics chain based on waste monitoring can be described using the following chart (Fig. 1).

2.1. Processing waste monitoring data

In this study, we consider the waste monitoring data as series of measured waste quantity (kg) in certain time points and geographical locations. In addition, the data could contain some related information (e.g. container, waste type) related to a measurement point. In the first stage, the raw monitoring data have to be processed to a format required by data analytics methods. Instead of discrete timestamp specific waste quantity, waste generation can be expressed in terms of continuous time-series depending on sampling resolution (kg per day, week, year, etc.). By combining weighing data with other data sources, more complicated indicators can be derived (see Fig. 1). For instance, it is possible to express waste generation in relation to external factors (e.g. kg per capita, kg per household, kg per inhabitant, kg per square meter, etc.).

Furthermore, various measures related to waste management can be estimated from the monitoring data. The fusion with data from TCS (e.g. route data) enables deriving various site or area specific transportation related indicators (e.g. distance per transported kg, MJ per transported kg, CO_2 per transported kg) if waste quantity, content and transportation distance are known. For

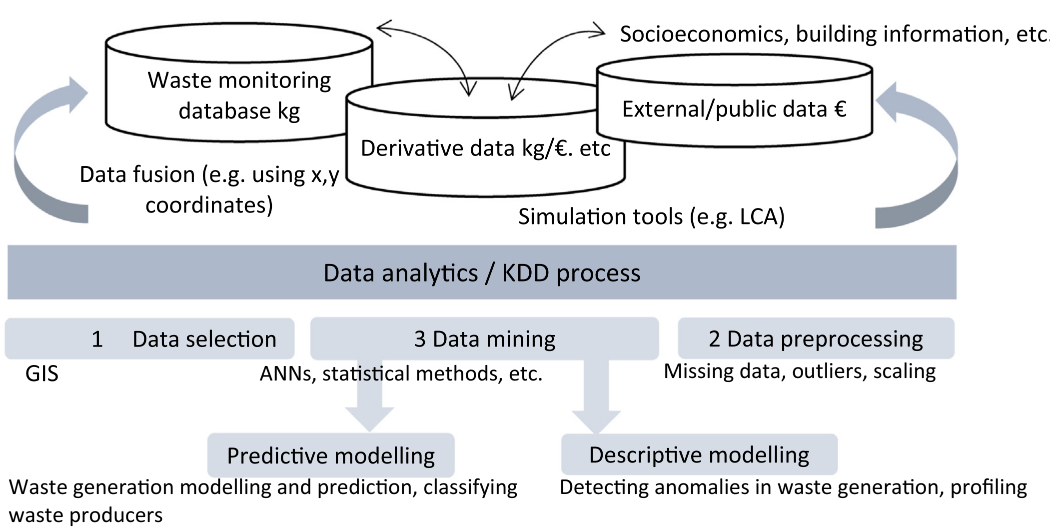

Fig. 1. The main outlines of waste monitoring data analytics.

Please cite this article in press as: Niska, H., Serkkola, A. Data analytics approach to create waste generation profiles for waste management and collection. Waste Management (2018), https://doi.org/10.1016/j.wasman.2018.04.033