Review of Data mining applications in forestry sector

Broz, Diego[1]; Olivera[2], Alejandro; Viana Céspedes, Víctor[2]; Rossit, Daniel Alejandro[3]

[1] UNaM CONICET Facultad de Ciencias Forestales
Bertoni 124, Eldorado (N3382GDD), Misiones, Argentina.
diego.broz@facfor.unam.edu.ar

[2] Universidad de la República
Ruta 5, km 386,5, 45000 Tacuarembó, Uruguay.
alejandro.olivera@cut.edu.uy, vviana@cci.edu.uy

[3] Departamento de Ingeniería, Universidad Nacional del Sur, CONICET
Av. Alem 1253, Bahía Blanca (B8000CPB), Argentina.
daniel.rossit@uns.edu.ar

Abstract

Modern technology makes possible to collect large amount of data that can be processed and transformed in valuable information for several human activities. Forest industry particularly can take advantage of such technology because of modern forest harvesters are equipped with a system for data collection and communication called StanForD. Data mining allows users to process large databases to determine trends and patterns. In this extended abstract we present a brief revision of the literature dedicated to the issue and, also, we indicate synthetically future research directions that could be useful for forest operations management. Some DM techniques are artificial neural network and decision tree and they are used to perform association, classification and clustering. Nonetheless, data mining techniques have been successfully applied to several fields, e.g. industry, marketing, sociology, economy, agriculture and environmental sciences.

1 Introduction

The amount of data recorded and stored daily is growing due to the use of technology with automatic data collection capabilities. By processing this data trends and patterns can be determined to use as input in the decision making process in many activities.
Given the amount of data available, to transform the data into information it is necessary to use special techniques that can handle and process the data. Data mining (DM) techniques arise as the solution to the problem. DM is the process of applying Computer Based Information Systems (CBIS) for discovering knowledge from data [1]. DM applications started in the 1960s, its baseline is grounded by disciplines such as machine learning, artificial intelligence, probability and statistics [2].
Knowing the impact variables on the forest harvesting productivity, transport of forest products, plantations establishment, silvicultural practices, among others factors, would make possible to manage such variables more efficiently across the forest industry. In addition, the technology used for harvesting in forest plantations in countries like Uruguay provides a mechanism to automatically record data from forest harvesters during the operation. In this sense, the DM allows to discover patterns and trends useful to predict the behavior of a system and interrelations of interest [3]. Thus, it would be a strategical issue to process that automatically collected data in order to support the decision making process. In this extended abstract we present a brief literature review of data mining implementations on forestry systems, and then, some new possible implementations with impact in forestry management are discussed.

2 Literature review

Different techniques are used for DM, e.g. artificial neural network, regression trees and decision tree (DT), and others [2, 4]. DM has been successfully applied to industrial processes manufacturing, marketing, sociology and others; however, there is little evidence of application to forestry, environmental sciences and agriculture. In [5], authors compared linear and regression tree analyses for

forest attribute estimations and their spatial modeling. The results of analysis showed that, statistical models of stand volume, tree density, species richness and reciprocal of Simpson indices using tree regression analysis had higher adjusted compared to linear regression models. Using DM techniques, [6] estimated the risk on forest fire and some of the methods analyzed were multilayer perceptron, radial basis function networks, support vector machines and fuzzy logic. They used historical forest fire records, which contained parameters like geographical conditions of the existing environment, date and time when the fire broke out, meteorological data such as temperature, humidity and wind speed, and the type and tree stocking. In [7], the authors used a DM methodology named instance-based classification for estimating carbon storage in *Araucaria angustifolia* (Bertol.) Kuntze plantation in Brazil. They concluded that the technique outperformed the conventional methods.

In [8], authors examined and analyzed three European projects as guidance to describe current possibilities and future challenges for deployment of Big Data (BD) techniques in the field of agro-environmental research, facilitating decision support at the level of societal challenges. The authors recommended the use of BD to analyse data from various sources, e.g. harvesting, production, and meteorological records. [9] were pioneers on integrating GNSS with forest harvester data to improve forest management. They retrieved data from a GNSS-enabled harvester working in cut-to-length operations in *Eucalyptus spp*. plantations in Uruguay. The dataset obtained comprised over 63,000 cycles of felled and processed stems. With this data, a mixed effects model was fitted to evaluate harvester productivity as a function of stem diameter at breast height, species, shift, slope and operator. To analyze the relevant economic, social and ecological factors of China's forestry resources [10] used the BD theory. Firstly, the authors used the method of data envelopment analysis to investigate the forestry resources efficiency; then they analyzed time series data using the Malmquist total factor productivity index method. Applying Neural network-based models [11] presented a large-scale evaluation of climate effects on the productivity of three temperate tree species in Central Europe. Using this technique they determined which among 13 tested climate variables best predicted the tree species-specific site index. To the best of our knowledge there is no evidence of studies using DM techniques applied to forest harvesting operations or data automatically collected by harvesting machines. As such, there is a potential field of application of DM techniques and compare the results against conventional Regression analysis as performed [9].

3. Conclusion

Various DM techniques have been applied in research for the agro-environmental sciences, including forestry. Prediction of forest fires, the effect of climatic variable on forest productivity, forests structure analysis and carbon storage are some of the case studies published. Techniques comprise mixed models, artificial neural network, association rules, and regression trees. However, there is still a gap regarding the use of DM techniques in forest operations, concretely using the automatic data collection system available in the majority of modern forest harvesters. This data enables to describe internal processes of the system based on actual data. Following these ideas, interesting future research could be oriented to estimates internal parameters that have significant impact on forestry operations planning, such as productivity rate (volume of harvested wood per hour, for example) and operations time (processing and transport times). In a more strategic management perspective, new maps can be developed. Analysing data from past harvesting campaigns, new fit-for-purpose forest yields maps can be built. Also, it is possible to assess and redefine internal forest roads according to the real land and site yield.

References

[1] Vlahos, G. E., Ferratt, T. W., Knoepfle, G. The use of computer-based information systems by German managers to support decision making. *Information & Management*, 41(6), 763-779, 2004.

[2] Jothi, N., Husain, W. Data mining in healthcare–a review. *Procedia Computer Science*, 72, 306-

313, 2015.

[3] Ahlemeyer-Stubbe, A., Coleman, S. *A practical guide to data mining for business and industry*, John Wiley & Sons, 2014.

[4] Liao, S. H., Chu, P. H., Hsiao, P. Y. Data mining techniques and applications–A decade review from 2000 to 2011. *Expert systems with applications*, 39(12), 11303-11311, 2012.

[5] Mohammadi, J., Shataee, S., Babanezhad, M. Estimation of forest stand volume, tree density and biodiversity using Landsat ETM+ Data, comparison of linear and regression tree analyses. *Procedia Environmental Sciences*, 7, 299-304, 2011.

[6] Özbayoğlu, A. M., Bozer, R. Estimation of the burned area in forest fires using computational intelligence techniques. *Procedia Computer Science*, 12, 282-287, 2012.

[7] Sanquetta, C. R., Wojciechowski, J., Dalla Corte, A. P., Rodrigues, A. L., Maas, G. C. B. On the use of data mining for estimating carbon storage in the trees. *Carbon balance and management*, 8(1), 1-9, 2013.

[8] Lokers, R., Knapen, R., Janssen, S., van Randen, Y., Jansen, J. Analysis of Big Data technologies for use in agro-environmental science. *Environmental Modelling & Software*, 84, 494-504, 2016.

[9] Olivera Farias, A. *Exploring opportunities for the integration of GNSS with forest harvester data to improve forest management*. PhD thesis, University of Canterbury, Nueva Zelanda, 2016.

[10] Li, L., Hao, T., Chi, T. Evaluation on China's forestry resources efficiency based on big data. *Journal of Cleaner Production*, 142, 513-523, 2017.

[11] Hlásny, T., Trombik, J., Bošeľa, M., Merganič, J., Marušák, R., Šebeň, V., Štěpánek, P., Kubišta, J., Trnka, M. Climatic drivers of forest productivity in Central Europe. *Agricultural and Forest Meteorology*, 234, 258-273, 2017.