Predictive models of forest development in Slovenia

Marko Debeljak [1*], Bernard Ženko [1], Aleš Poljanec [2]

[1] Jožef Stefan Institute, Department of knowledge technologies

[2] University of Ljubljano, Biotechnical Faculty

[*] Jamova cesta 39, 1000 Ljubljana, Slovenia, tel.: 01 4773124, e-mail: marko.debeljak@ijs.si

Introduction

In the area of ecological data analysis there is an increasing demand for methods and tools based on novel approaches from machine learning and information theory that would complement classic statistical methods. This would significantly increase the number of tasks that can be addressed with data analysis and provide higher quality of the analysis results. Data mining, for example, uses machine learning methods that employ approaches from classical statistics as well as information theory. Machine learning tools have been successfully used for data analysis and learning of qualitative and quantitative models from the data. Models can be written in a human readable form (e.g., decision rules and trees, equations) or in a form that can only be used for predicting new examples (e.g., neural networks, support vector machines, etc.). For the purpose of the analysis of ecological data, decision trees are frequently the method of choice. Due to their hierarchical structure, models in the form of decision trees are easy to interpret and can be used to predict values of the target variable that can be simple or structured (e.g., a vector, a hierarchy, etc.). We present a pilot study of applying data mining techniques in order to analyze the forest in Slovenia and to develop scenarios of its future development that could be used in forest management. We have focused on the analysis of the time dynamics of wood stock changes and developed models that enable us to investigate the factors that influence the total wood stock in a forest, as well as to predict the wood stock in the future.

Methods

The basis of our study was the *Silva-SI* database (Poljanec, 2008). The database comprises data from 21052 compartments in five periods from year 1970 to 2008 (1970, 1980, 1990, 2000 and 2008). Depending on the share of state owned forest in a given compartment, we have formed two groups: "state owned" forest, which is 100% owned by the state (5237 compartments), and "private owned" forest, where the share of state owned forest is below 100% (15815 compartments). We have modeled the dynamics of wood stock changes within the compartments with predictive clustering trees (Blockeel et al., 1998) implemented in the Clus data mining system (Blockeel & Struyf, 2002). Models for predicting the total wood stock in each forest compartment in year 2018 were learned with model trees in the Weka data mining suite (Witten & Frank 2005). Besides the models for the total wood stock, we have also built separate models for different thickness classes (A, B and C), but due to space constraints, we at this point only present the models for the total wood stock in "private" forests.

Results

The structure of the model describing the dynamics of wood stock changes within the compartments (Fig. 1) shows that the most important attributes are phytogeographic region and growing region coefficient followed by substratum and altitude. The trends of wood stock changes are all increasing. The model evaluation shows that going from the first period (1970) to the present the errors (absolute mean error, root mean squared error) are decreasing, while the correlation coefficient remains roughly the same through all periods.

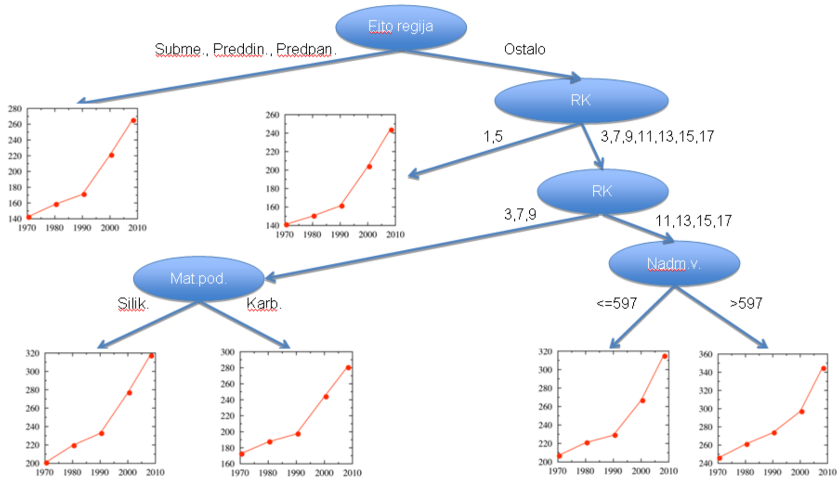

Fig. 1: The model describing dynamics of wood stock changes in "private" forests.

Descriptive models for total wood stock in different forest compartments for specific periods from year 1970 to 2008 show that the single most important factor affecting wood stock in the selected period is the wood stock in the preceding period. Given the absence of data before year 1970, the model of wood stock in year 1970 does not include the attribute for the "wood stock in the preceding period"; the most important attribute in this case is phytogeographic region. Evaluation of this model, however, gives significantly lower correlation coefficient and suggests that the model is of a limited value.

The amount of the total wood stock in year 2018 was assessed by extrapolating the trends from the leaves of the model tree learned for the year 2008 (Fig. 2). We have validated this methodology by extrapolating the model for the wood stock in year 2000 to the year 2008 and compared this predictions with the real data for year 2008. The differences between the predicted and real values were small (the largest difference was 9.2%), so we believe this approach can give us sound estimates of the wood stock in the future.

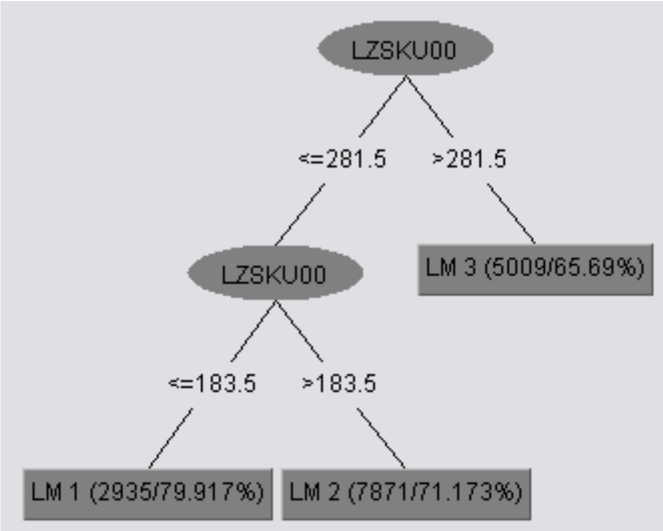

Fig. 1: The model describing wood stock in "private" forests in year 2008.

The coefficients of extrapolated linear models LM1, LM2 and LM3 for year 2018 suggest further increase of wood stock, provided that the forest management will not significantly change in the next decade. The largest increase can be expected in compartments that had low wood stock in 2008, while the smallest increase will be in compartments with already high wood stock (Table 1).

Table 1: Predicted wood stock for "private" forest in year 2018.

	LM1	LM2	LM3
Extrapolation function coefficients; y=kx+n (k, n)	9,48; - 18815	5,84; -11464	1,83; -3336,6
Average true wood stock in year 2008	224,4	280,0	351,7
Predicted wood stock in year 2008	224,5	279,6	351,7
Predicted wood stock in year 2018	319,3	338,1	370,1
Growth index (year 2008 =100)	142,3	120,8	105,2

Zaključek

Na osnovi dobljenih rezultatov lahko zaključimo, da podatkovna zbirka *Silva-SI* omogoča zanesljivo uporabo metod podatkovnega rudarjenja za izdelavo modelov časovne dinamike lesne zaloge, modelov za razlago stanja lesne zaloge po posameznih obdobjih in za izgradnjo napovednih modelov lesne zaloge. Uporabljen pristop omogoča dinamično modeliranje, ki optimizira strukturo modela glede na značilnosti podatkov, omogoča oblikovanje podskupin podatkov in določitev najvplivnejših dejavnikov in povezave med njimi (odkrivanje direktnih in indirektnih vplivov) in tako prispeva k pridobivanju novega znanja iz podatkov. Z uporabo kvalitativnega odločitvenega modeliranja bi rezultate podatkovnega rudarjenja lahko povezali z obstoječim ekspertnim znanjem ter gozdarstvu ponudili učinkovit in uporabniku prijazen sistem za podporo odločanju o prihodnjem ravnanju z gozdovi.

Literatura

H. Blockeel, L. De Raedt, and J. Ramon 1998. Top-down induction of clustering trees, Proceedings of the 15th International Conference on Machine Learning (Shavlik, J., ed.), pp. 55-63,

Blockeel, H., Struyf, J. 2002. Efficient algorithms for decision tree cross-validation. Journal of Machine Learning Research, 3:621–650.

Poljanec, A. 2008. Strukturne spremembe gozdnih sestojev v Sloveniji v obdobju 1970-2005. Doktorska disertacija, Ljubljana: 126 s.

Witten, I.H., Frank, E. 2005. Data Mining: Practical Machine Learning Tools and Techniques. Morgan Kaufmann, San Francisco, CA, USA, second edition.