Random Forest Machine Learning Technique for Automatic Vegetation Detection and Modelling in LiDAR Data

Fayez Tarsha Kurdi[1]*, Wijdan Amakhchan[2] and Zahra Gharineiat[3]

[1]*Institute of Integrated and Intelligent Systems, Griffith University, Australia*

[2]*Faculté des Sciences et Techniques de Tanger*

[3]*School of Civil Engineering & Surveying, Faculty of Health, Engineering and Sciences, University of Southern Queensland, Australia*

*Corresponding author: Fayez Tarsha Kurdi, Research Fellow, Institute of Integrated and Intelligent Systems, Griffith University, Nathan, QLD 4111, Australia

Abstract

Machine learning techniques have gained a distinguished position in the automatic processing of Light Detection and Ranging (LiDAR) data area. They represent the actual research topic in the remote sensing domain. Indeed, this paper presents one method of supervised machine learning, which is called Random Forest. This algorithm is discussed, and their primary applications in automatic vegetation extraction and modelling in the LiDAR data area are presented here.

Keywords: LiDAR; Random forest; Classification; Modelling

Introduction

Nowadays, Light Detection and Ranging (LiDAR) data wins an advanced position among other remote sensing data [1]. Automatic vegetation detection and modelling in forest and urban areas are one of the important envisaged applications of LiDAR data. In fact, automatic tree detection in LiDAR data belongs to the automatic classification of LiDAR data topic. The scanned scene consists of different man-made objects such as buildings, bridges, roads and dams. Furthermore, the studied zone may contain natural item classes such as vegetation, terrain, rivers, and lakes. In order to model the project area, its LiDAR point cloud has to be classified according to the main classes [2]. Once the classification is achieved successfully, the next step is to model each class aside. Concerning the vegetation detection and modelling, classic approaches were employed in the literature such as RANdom Sample Consensus (RANSAC) algorithm [3,4], local maximum algorithm [5], surface growing algorithm and multiple echo analysis [6], voxel layer single tree modelling algorithm [7], morphological algorithm [8] and analysis of fullwave form LiDAR data [9].

Recently, a modern technique called machine learning enhanced the automatic processing of the LIDAR data area. This technique becomes quickly widespread and occupied a major position in regard to the other classical approaches. This paper presents one machine learning method that widely applied for automatic vegetation recognition and modelling in LiDAR data field. This method is Random Forest (RF). The paper aims to summarize the principal of this technique in addition to its main applications in LiDAR data field.

In the next section, Random forest technique will be discussed.

Random Forest

RF is an ensemble of supervised learning algorithms used for classification and regression, used in predictive modelling and machine learning technique [10]. It gathers the results and the predictions of several decision trees to finally choose the best output which is the mode (the value that appears most often in the set of decision trees results) of the classes or mean prediction.

RF works by splitting the dataset into two sections, the training set and the test set. Then randomly select multiple samples from the training set. Next, use the decision tree for each sample which divides each selection into two daughters using the best division. Thereafter, repeat the last step to finally vote for each prediction result and select the most voted prediction as the final result (Figure 1).
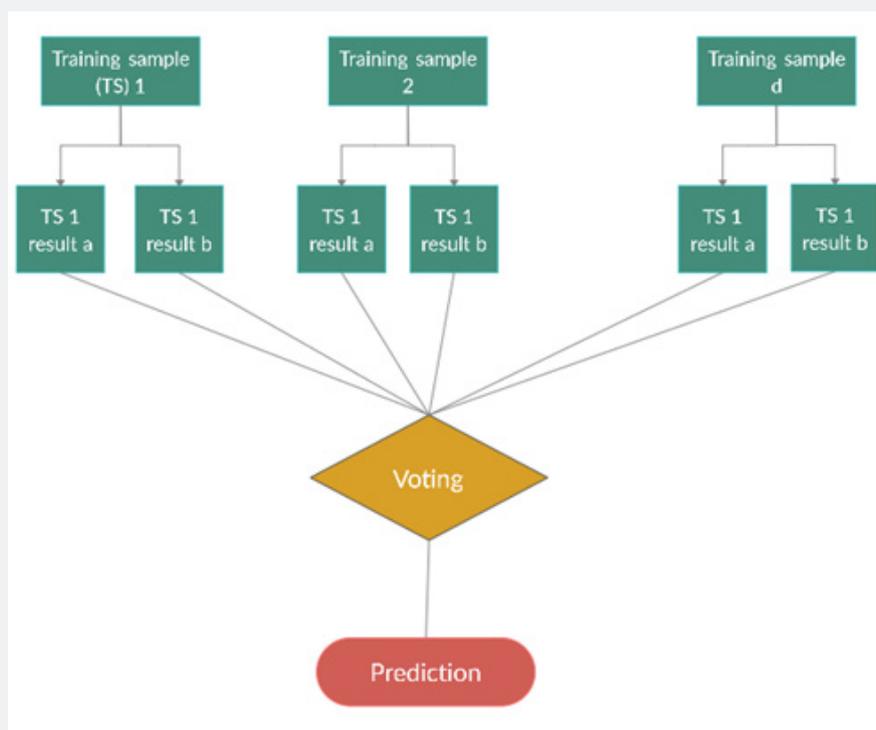
Figure 1: Random Forest training workflow.

The main hyperparameters in Random Forest are either used to increase the predictive power of the model or to make the model faster [11]. In this context, a higher number of trees can increase the performance as well as makes the predictions more stable, but unfortunately, the processing time becomes longer. Furthermore, the employment of a maximum number of features in addition to a minimum number of leaves are that requested splitting internal nodes may improve the algorithm performance. Once the training step is realized, the trained model can be applied to a dataset that is not used for training. This procedure allows estimating their predictions and compared them to the expected values [12].

In literature, many authors applied RF exclusively on LiDAR data [13,14] whereas other authors simultaneously used additional data and LiDAR point cloud as input for RF algorithm [15,16]. From another viewpoints, several applications were achieved on LiDAR data by using the RF technique. Yu et al. [14] suggested an approach for estimating tree characteristics such as height, diameter, and stem volume using LiDAR data. For attending this goal, RF is considered as a classifier. Levick et al. [17] fused the Digital Surface Model (DSM) calculated from LiDAR point cloud and field-measured wood volume using RF Algorithm. Chen et al. [13] used the feature selection method and RF algorithm for forested landslide detection. For this purpose, the Digital Terrain Model (DTM) and the slope model was established for the scanned scene, and the selected features are calculated at the pixel level. The same principle was applied by Guan et al. [18] to classify the city components in the urban zones.

RF was broadly used for vegetation detection in forest and urban areas. Niemeyer et al. [19] classified the scanned city elements by integrating RF classifier into a Conditional Random Field (CRF) framework. Moreover, Man et al. [16] extracted grasses and trees in urban areas using airborne LiDAR and hyperspectral data. RF and object-based classification methods were employed together to extract the distribution map of urban vegetation. [20,21] underlined the efficiency of RF for vegetation detection in forest and urban areas. Huang & Zhu [15] developed an approach for fusing hyperspectral image and LiDAR data based on RF. In this context, each feature is ranked by RF, and more useful features are selected as inputs for RF for data classification.

Conclusion

RF is an efficient machine learning technique that can be used for automatic vegetation extraction and modelling in forests and urban zones in LiDAR data. In this context, LiDAR data can be used exclusively or in addition to other supplementary data such as field-measurement and hyperspectral data.

References

1. Tarsha Kurdi F, Awrangjeb M (2020) Comparison of LiDAR building point cloud with reference model for deep comprehension of cloud structure. Canadian Journal of Remote Sensing 46(5): 603-621.

2. Tarsha Kurdi F, Awrangjeb M, Munir N (2021) Automatic filtering and 2D modelling of LiDAR building point cloud. Transactions in GIS Journal 25(1): 164-188.

3. Burt A, Disney, M, Calders K (2018) Extracting individual trees from LiDAR point clouds using treeseg. Methods in Ecology and Evolution 10(3): 438-445.

4. Monnier F, Vallet B, Soheilian B (2012) Trees detection from laser point clouds acquired in dense urban areas by a mobile mapping system. ISPRS Annals of Photogrammetry, Remote Sensing and Spatial Information Sciences I–3: 245-250.

5. Zhang C, Zhou Y, Qiu F (2015) Individual Tree Segmentation from LiDAR Point Clouds for Urban Forest Inventory. Remote Sensing 7(6): 7892-7913.

6. Liu J, Shen J, Zhao R, Xu S (2013) Extraction of individual tree crowns from airborne LiDAR data in human settlements. Mathematical and Computer Modelling 58(3-4): 524-535.

7. Wang Y, Weinacker H, Koch B, Sterenczak K (2008) LiDAR point cloud based fully automatic 3D single tree modelling in forest and evaluations of the procedure. Int Archives Photogrammetry Remote Sens Spatial Inform Sci XXXVII: 45-51.

8. Vauhkonen J, Ene L, Gupta S, Heinzel J, Holmgren J, et al. (2012) Comparative testing of single-tree detection algorithms under different types of forest. Forestry 85(1): 27-40.

9. Gupta S, Weinacker H, Koch B (2010) Comparative Analysis of Clustering-Based Approaches for 3-D Single Tree Detection Using Airborne Fullwave LiDAR Data. Remote Sensing 2(4): 968-989.

10. Breiman L (2001) Random Forests. Machine Learning 45(1): 5-32.

11. Hastie T, Tibshirani R, Friedman J (2009) The Elements of Statistical Learning: Data Mining, Inference, and Prediction. (2nd edn), Springer Science & Business Media.

12. Brownlee J (2020) Train-Test Split for Evaluating Machine Learning Algorithms. Machine Learning Mastery.

13. Chen W, Li X, Wang Y, Chen G, Liu S (2014) Forested landslide detection using LiDAR data and the random forest algorithm: A case study of the three Gorges, China. Remote Sensing of Environment 152: 291-301.

14. Yu X, Hyyppä J, Vastaranta M, Holopainen M, Viitala R (2011) Predicting individual tree attributes from airborne laser point clouds based on the random forests technique. ISPRS Journal of Photogrammetry and Remote Sensing 66(1): 28-37.

15. Huang R, Zhu J (2013) Using Random Forest to integrate LiDAR data and hyperspectral imagery for land cover classification. 2013 IEEE International Geoscience and Remote Sensing Symposium - IGARSS 2013: 3978-3981.

16. Man Q, Dong P, Yang X, Wu Q, Han R (2020) Automatic Extraction of Grasses and Individual Trees in Urban Areas Based on Airborne Hyperspectral and LiDAR Data. Remote Sensing 12(17): 2725.

17. Levick SR, Hessenmöller D, Schulze ED (2016) Scaling wood volume estimates from inventory plots to landscapes with airborne LiDAR in temperate deciduous forest. Carbon Balance Manage 11(7).

18. Guan H, Yu J, Li J, Luo L (2012) Random forests-based feature selection for land-use classification using LiDAR data and orthoimagery. International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences, Volume XXXIX-B7, 2012 XXII ISPRS Congress, 25 August – 01 September 2012, Melbourne, Australia.

19. Niemeyer J, Rottensteiner F, Soergel U (2013) Classification of urban LiDAR data using conditional random field and random forests. Joint Urban Remote Sensing Event 2013: 139-142.

20. Michałowska M, Rapinski J (2021) A Review of Tree Species Classification Based on Airborne LiDAR Data and Applied Classifiers Remote Sens 13(3): 353.

21. Chehata N, Guo L, Mallet C (2009) Airborne LIDAR feature selection for urban classification using random forests. International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences, p. 38.

Your next submission with Juniper Publishers will reach you the below assets

- Quality Editorial service
- Swift Peer Review
- Reprints availability
- E-prints Service
- Manuscript Podcast for convenient understanding
- Global attainment for your research
- Manuscript accessibility in different formats
  (Pdf, E-pub, Full Text, Audio)
- Unceasing customer service

Track the below URL for one-step submission
https://juniperpublishers.com/online-submission.php