

Thesis

Improving models for environmental applications of LiDAR: Novel approaches based on soft computing

Jorge Garcia-Gutierrez

Department of Computer Languages and Systems, University of Seville, Seville, Spain

E-mail: jorgarcia@us.es

Abstract. This work proposes novel methodologies to improve the use of Light Detection And Ranging (LiDAR) for environmental purposes, especially for thematic mapping (LiDAR only or fused with other remote sensors) and the estimation of forest variables. The methodologies make use of well-known techniques from soft computing (machine learning and evolutionary computation) and their adaptation to develop LiDAR-derived products.

Keywords: LiDAR, machine learning, remote sensing, soft computing

1. Introduction

Light Detection And Ranging (LiDAR) is a remote sensing laser-based technology which can determine the distance to an object based on the round-trip time of an emitted pulse. LiDAR provides not only the x - y position but also the coordinate z for every impact, besides the information of the intensity of the signal reflected. LiDAR's suitability for the extraction of vertical information has widespread use in geomatics, archaeology, geography, geology, geomorphology, seismology, forestry, remote sensing and atmospheric physics [1].

Although LiDAR has become a popular sensor, most advanced modelling techniques have not been explored to improve on traditional techniques such as stepwise regression. Thus, we propose two novel methodologies for classification and regression on LiDAR based on the concept of ensembles of machine learning techniques and evolutionary computation (EC). A thorough description can be found in the original research work [2].

The first relevant contribution is devoted to avoid problems related to pixel-based classification without the use of complex segmentation procedures. Support

Vector Machines (SVMs) among other techniques such as decision trees [3] have been proved to be suitable for LiDAR classification. We propose a two-step classification by means of a stacking of a support vector machine and a refinement for every pixel by the application of the nearest-neighbour rule in its 8-adjacency.

The second main contribution consists in improving the selection of LiDAR-derived statistics in regression models for the estimation of forest variables such as dominant height or above-ground biomass. The traditional stepwise regression has some well-known drawbacks which evolutionary feature selection avoids.

Finally, our work aims to apply the proposed methodologies to different areas of high relevance in the Iberian peninsula in order to statistically validate the proposed methods' best performance.

2. Methodology

2.1. LiDAR classification

We propose a methodology [5] based on the use of a SVM to classify LiDAR fused with other data sources and a posterior refinement of every pixel by the use of

the nearest-neighbour rule. Features in the second level are weighted. Each feature is assigned a set of weights for each different label in the training phase. Later, an instance can select the subset of weights according to its most probable label [6]. Furthermore, the search of the most appropriate weights is carried out by means of EC with a fitness based on the optimization of the final accuracy in the training phase.

2.2. LiDAR regression

For regression, we propose a hybrid approach based on joint use of the common-in-literature multiple linear regression (proved valid in a small-size context such as plot-level datasets) and a genetic algorithm for the selection of the best LiDAR statistics [4] in the context of the estimation of forest variables. Genetic selection provides an affordable solution in terms of computational cost in contrast with an exhaustive search where the increasing number of statistics may be an important problem and the classical stepwise regression. The genetic algorithm defined is a common genetic selection but its fitness function tries to optimize the Bayesian Information Criterion and avoid multicollinearity problems.

3. Results

The proposed methodologies were assessed in data from several areas in the Iberian peninsula. The results obtained have been compared with other existing techniques in terms of quality measures.

Specifically, real data from a riparian area in southern Spain ceded by the Regional Ministry of Andalusia were processed to develop a thematic map. The results showed that our classification method improved the results of SVMs with a global accuracy over 90%.

Moreover, several experiments were performed to compare the different feature selection techniques in the context of variables estimation from LiDAR. Data from two areas in the north of Spain were used to evaluate the different methodologies. The obtained results showed that our genetic proposal outperformed stepwise and best-subset approaches in terms of BIC.

Conclusions obtained from the results in both experiments were statistically confirmed by the use of the corresponding non-parametric statistical tests.

4. Conclusions

In this work, a methodology based on soft computing techniques for the extraction of knowledge from LiDAR data was proposed. More specifically, we presented a hybrid classification technique and an evolutionary scheme to enhance the effectiveness of traditional selection of LiDAR-derived statistics for the estimation of forest variables. The proposed methodologies were tested on LiDAR data from several areas of the Iberian peninsula and their suitability was confirmed after providing high-quality results. Additionally, our methodology statistically outperformed other existing techniques.

Currently, we are addressing the problems related to the small size of the available data for both classification and regression tasks. Collecting additional data is problematic due to the excessive costs associated to their generation. Instance generation techniques may be a tool to overcome this limitation.

Acknowledgements

The research was supported by the Spanish Research under projects TIN2007-68084-C-02-01 and TIN2011-28956-C02-02 and the Andalusian Research project TIC-7528.

References

- [1] M.N. Avadhanulu, *An Introduction to Lasers Theory and Applications*, S. Chand Limited, 2001.
- [2] J. Garcia-Gutierrez, Intelligent techniques on LiDAR for environmental applications, PhD thesis, Escuela Técnica de Ingeniería Informática, University of Seville, June 2012.
- [3] J. Garcia-Gutierrez, L. Goncalves-Seco and J.C. Riquelme-Santos, Automatic environmental quality assessment for mixed-land zones using lidar and intelligent techniques, *Expert Syst. Appl.* **38**(6) (2011), 6805–6813.
- [4] J. Garcia-Gutierrez, E. Gonzalez-Ferreiro, J.C. Riquelme-Santos, D. Miranda, U. Dieguez-Aranda and R.M. Navarro-Cerrillo, Evolutionary feature selection to estimate forest stand variables using lidar, *Int. J. Applied Earth Observation and Geoinformation* **26** (2014), 119–131.
- [5] J. Garcia-Gutierrez, D. Mateos-Garcia and J.C. Riquelme-Santos, EVOR-STACK: A label-dependent evolutive stacking on remote sensing data fusion, *Neurocomputing* **75**(1) (2012), 115–122.
- [6] D. Mateos-Garcia, J. Garcia-Gutierrez and J.C. Riquelme-Santos, On the evolutionary optimization of k-NN by label-dependent feature weighting, *Pattern Recognition Letters* **33**(16) (2012), 2232–2238.