This is an excerpt from the thesis "*Ensemble Relearning for Building Type Classification with Remote Sensing Data*".

Please contact Fernanda Abigail Bosmediano Chiquin for a full version of the thesis.

Hochschule
für Technik
Stuttgart
University of Applied Sciences

Master of Science Programme
Photogrammetry and Geoinformatics
Master Thesis
Winter Term 2018/2019

Ensemble Relearning for Building Type Classification with Remote Sensing Data

by

Fernanda Abigail Bosmediano Chiquin

Supervisors: **Prof. Dr.-Ing. Michael Hahn** (Hochschule für Technik Stuttgart)
Dr. Christian Geiß (Deutsches Zentrum für Luft- und Raumfahrt)

Master Course Photogrammetry and Geoinformatics

Ensemble Relearning for Building Type Classification with Remote Sensing Data

Abstract

Building type classification is a critical element in building inventories, which are essential in earthquake losses estimation. The collapse of buildings mainly causes the death toll related to earthquakes. However, such inventories are frequently not available or are incomplete. To compile the required building inventory data and assign relevant features to the buildings often includes detailed building-by-building assessments which require ample time and financial investment. To overcome these obstacles, remote sensing techniques have shown the potential to extract relevant features for characterization of buildings and subsequent vulnerability analysis. This study introduces a learning method for assigning the building type to a building inventory using features from remote sensing data and limited in situ observations. The method achieved an overall accuracy of 76.01% and built upon an ensemble of supplementary machine learning algorithms and techniques such as Random Forest, Nearest Neighbor, Gradient Boosting and Stacking learning. In the second stage, a new method to increase the accuracy of the model is proposed. The selected model was applied to a sample of 20,000 buildings. An accuracy of 72. 32% was reached. The prediction of this model has been added as a new feature and has been a model relearned. With this prediction, three new features have been calculated using the majority filter concept. The model was relearned for a second time, and an accuracy of 72.75% was attained.

Keywords: Seismic building structural type; Machine Learning; Ensemble Learning; Relearning Process; Cologne; Earthquake

Table of Contents

Table of Figures

Table of Tables

Table of Appendices

Abbreviations

AB	AdaBoost
DT	Decision Tree
GB	Gradient Boosting
GNB	Gaussian Naïve Bayes
KNN	K Nearest Neighbor
LDA	Linear Discriminant Analysis
LR	Logistic Regression
RF	Random Forest
SVM	Support Vector Machine

1. Introduction

Building type classification is a vital element in building inventories, which are essential in earthquake loss estimation. In fact, in 2017, worldwide economic losses from disasters were assessed at $337 billion (Swiss Re, 2018). During the same year, a total of 92.80 million people were affected by natural disasters (Roser & Ritchie, 2018), with earthquakes causing approximately 1,012 deaths (Below & Wallemacq, 2018).

Earthquakes occur daily around the globe. However, the disaster risk of a system is probabilistically determined as a function of hazard, exposure, vulnerability, and capacity (United Nations Office for Disaster Risk Reduction, 2017). Nowadays, people are becoming more vulnerable to earthquakes regardless of whether they live in rich or emerging countries (Dan et al., 2014). Indeed, the death toll related to earthquakes is mainly caused by the collapse of buildings (United States Geological Survey, 2018). Therefore, it is clear that building inventory data requires reliable estimation of earthquake damage. However, some countries do not have enough data for such estimations, and, even if they have them, there is plenty of work to do (Matsuoka et al., 2014). Similarly, building inventory and its vulnerability, especially for earthquake losses estimation, usually involve a considerable amount of time and money (Dunbar et al., 2003).

There are different approaches to seismic vulnerability evaluation of existing buildings. The conventional methods are designed by structural engineers and require detailed assessments of each building. They are costly and sometimes unable to cope with large areas (Geiß et al., 2015). Instead, different remote sensing techniques have proven their potential to extract relevant features to assess earthquake risk (Geiß & Taubenböck, 2013). For instance, Geiß et al. (2015) combined scarce in situ observations, multisensory remote sensing data, and machine learning techniques to estimate seismic building structural types in the city of Padang (Indonesia). The study performed a supervised classification with the models that were built using Support Vector Machine (SVM) algorithm and Random Forest (RF) independently. It was found that one model performed better with some features than the other.

Likewise, a machine learning classification of buildings was completed by Lee et al. (2017). The project applied four different algorithms: Decision Tree (DT), K Nearest Neighbor (KNN), Gaussian Naïve Bayes (GNB), and SVM. However, the study concluded that a reinforcement of the model is needed and that the application of other learning models should be investigated.

A study done by Li et al. (2018) highlights the importance of machine learning techniques during earthquake relief. It evaluated the seismic waveform recorded by 16 seismological stations and determined the time that vertical and horizontal waves reached a seismograph station.

SVM, RF, and DT algorithms were applied individually. It was settled that a combination of some methods can be further analyzed and the addition of new features can improve the prediction results.

The literature discussed above shows the potential of machine learning for building classification and the importance of applying new methodologies during earthquake relief. They all agree that an application of other learning algorithms can be useful for improving the accuracy of the results. Therefore, this project intends to develop a new model to automatically classify buildings from the City of Cologne. This study combines different machine learning algorithms. Due to the time and cost of completing building inventories, this thesis proposes to increase the accuracy of the model with a computation of new features on the geospatial domain and relearning processes.

This is an excerpt from the thesis "*Ensemble Relearning for Building Type Classification with Remote Sensing Data*".

Please contact Fernanda Abigail Bosmediano Chiquin for a full version of the thesis.

References

Balakrishnama, S. & Ganapathiraju, A. (1998) *Linear Discriminant Analysis- a brief tutorial,* Mississippi State University, Department of Electrical and Computer Engineering, Institute for Signal and Information Processing, pp.1-2.

Below, R. & Wallemacq, P. (2018) *Annual Disaster Statistical Review 2017,* Centre for Research on the Epidemiology of Disasters, pp.4-5.

Bergstra, J. & Bengio, Y. (2012) *Random search for hyper-parameter optimization,* Journal of Machine Learning Research, MIT Press, Cambridge, Massachusetts, USA, 13(1), pp. 281-305.

Bergstra, J., Bardenet, R., Bengio, Y. & Kegl, B. (2011) *Algorithms for hyper-parameter optimization*, in: Shawe-Taylor, J., Zemel, R. S., Bartlett, P. L., Pereira, F. and Weinberger, K. Q. (Eds.), Advances in neural information processing systems, Curran Associates, New York, USA, pp. 2546-2554.

Breiman, L. (1996) *Bagging Predictors*, Machine Learning, Kluwer Academic Publishers, 24(2), pp. 123–140.

Brown, G. (2010) *Ensemble Learning,* in C. Sammut and G. I. Webb (Eds.), Encyclopedia of Machine Learning, Springer USA, Boston, Massachusetts, USA, pp. 312-320.

Brownlee, J. (2018) *Machine Learning Mastery with Python*. Melbourne, Australia.

Chen, T. & Guestrin, C. (2016) *XGBoost: A Scalable Tree Boosting System,* in Proceedings of the 22nd ACM SIGKDD Conference 2016, ACM, New York, USA, pp. 785-794.

Cover, H. & Hart, P. (1967) *Nearest neighbor pattern classification,* IEEE Transactions on Information Theory, IEEE, Piscataway, New Jersey, USA, 13(1), pp. 21-27.

Dan, M. B., Armas, J. & Goretti, A. (2014) *Earthquake Hazard Impact and Urban Planning,* Springer Netherlands, Dordrecht, Netherlands, pp.1-12.

Dunbar, P. K., Bilham, R. & Laituri, M. J. (2003) *Earthquake Loss Estimation for India Based on Macroeconomic Indicators,* in Beer, T., Ismail-Zadeh, A. (Eds.) Risk Science and Sustainability. NATO Science (Series II: Mathematics, Physics, and Chemistry), 112, pp. 163-180. Springer Netherlands, Dordrecht, Netherlands.

Freund, Y. & Schapire, R. (1996) *Experiments with a New Boosting Algorithm,* in Proceedings of the 13th International Conference on Machine Learning, Morgan Kaufmann Publishers, San Francisco, California, USA, pp. 148-156.

Friedman, J. H. (2002) *Stochastic gradient boosting,* Computational Statistics & Data Analysis, Elsevier, New York, USA, 38(4), pp. 367-378.

Geiß, C. & Taubenböck, H. (2013) *Remote sensing contributing to assess earthquake risk: from a literature review towards a roadmap,* Natural Hazards, Springer Netherlands, Dordrecht, Netherlands, 68(1), pp. 7-48.

Geiß, C. & Taubenböck, H. (2015) *Object-Based Postclassification Relearning,* IEEE Geoscience, and Remote Sensing Letters, IEEE, Piscataway, New Jersey, USA, 12(11), pp. 2336-2340.

Geiß, C., Jilge, M., Lakes, T. & Taubenböck, H. (2016) *Estimation of seismic vulnerability levels of urban structures with multisensor remote sensing,* IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, IEEE, Piscataway, New Jersey, USA, 9(5), pp. 1913-1936.

Geiß, C., Thoma, M., Pittore, M., Wieland, M., Dech S. W. & Taubenböck H. (2017) *Multitask active learning for characterization of built environments with multisensor earth observation data,* IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, IEEE, Piscataway, New Jersey, USA, 10(12), pp. 5583-5597.

Güneş, F., Wolfinger, R. & Tan, P. (2017) *Stacked Ensemble Models for Improved Prediction Accuracy*, in Proceedings of the SAS Global Forum 2017, SAS Institute, Cary, North Carolina, USA, pp. 1-5.

Koch, K. (1990) *Bayes' theorem,* Bayesian Inference with Geodetic Applications, Springer Germany, Berlin / Heidelberg, pp. 4-8.

Kohavi, R. (1995) *A study of cross-validation and bootstrap for accuracy estimation and model selection,* in Proceedings of the 14th IJCAI Conference, Morgan Kaufmann Publishers, San Francisco, California, USA, 2(14), pp. 1137-1145.

Lee, J., Jang, H., Yang, J. & Yu, K. (2017) *Machine Learning Classification of Buildings for Map Generalization,* ISPRS International Journal of Geo-Information, MDPI, Basel, Switzerland, 6(10), 309, pp. 1-15.

Li, W., Narvekar, N., Nakshatra, N., Raut, N., Sirkeci B. & Gao J. (2018) *Seismic Data Classification using Machine Learning,* in: IEEE Fourth International Conference on Big Data Computing Service and Applications (BigDataService), Proceedings, IEEE, Piscataway, New Jersey, USA, pp. 56-63.

Liaw, A. & Wiener, M. (2002) *Classification and regression by RF,* R news, 2(3), pp. 18-19.

Matsuoka, M., Mito, S., Midorikawa, S., Miura, H., Quiroz, L., Maruyama, Y., Estrada, M. (2014) *Development of building inventory data and earthquake damage estimation in Lima, Peru for future earthquakes,* Journal of Disaster Research, 9(6), pp. 1032-1041.

Mather, P. M. & Koch, M. (2011) *Computer Processing of Remotely/ Sensed Images: An Introduction,* John Wiley & Sons, pp. 267-270.

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot & M.Duchesnay, É. (2011) *Scikit-learn: Machine Learning in Python,* Journal of Machine Learning Research, MIT Press, Cambridge, Massachusetts, USA, 12(2/1/2011), pp. 2825–2830.

Ranawana, R. & Palade, V. (2006, April 24) *Multi-Classifier Systems: Review and a Road Map for Developers,* International Journal of Hybrid Intelligent Systems, 3(1), pp. 35-61.

Raschka, S. (2018) *StackingCVClassifier,* https://rasbt.github.io/mlxtend/user_guide/classifier/StackingCVClassifier/#methods (2nd October 2018)

Roser, M. & Ritchie, H. (2018) *Natural Catastrophes,* https://ourworldindata.org/natural-catastrophes (19 September 2018)

Schapire, R. & Freund, Y. (2012) *Boosting Foundations and Algorithms,* MIT Press, Cambridge, Massachusetts, USA, pp.17-25.

Sebastian, R. & Vahid, M. (2017) *Python Machine Learning Second Edition,* Packt Publishing, Birmingham, United Kingdom, pp.52-105.

Shai, S. & Shai, D. (2014) *Understanding Machine Learning From Theory to Algorithms,* Cambridge University Press, Cambridge, United Kingdom, pp. 5-7.

Suykens, J. & Vandewalle, J. (1999) *Least squares Support Vector Machine classifiers,* Neural processing letters, Springer US, 9(3), pp. 293-300.

Swiss Re Group (2018) *At USD 144 billion, global insured losses from disaster events in 2017 were the highest ever; sigma study says,* https://www.swissre.com/media/news-releases/2018/nr20180410_sigma_global_insured_loses_highest_ever.html (19 September 2018)

Tabachnick, B. G. & Fidell, L. S. (2007) *Using multivariate statistics 5^th Edition,* Allyn & Bacon, Needham Heights, Massachusetts, USA, pp. 437-481.

Balch, T. & Chakraborty, A. (2016) *Machine Learning for Trading,* UDACITY. Georgia Tech University.

United Nations Office for Disaster Risk Reduction (UNISDR) (2017) *Terminology,* https://www.preventionweb.net/terminology/view/7818 (19 September 2018)

United States Geological Survey (USGS) (2018) *Hazard and Risk Assessment,* https://earthquake.usgs.gov/research/hazrisk/risk.php (19 September 2018).

Weinberger, K. Q., Blitzer, J. & Saul, L. K. (2006) *Distance metric learning for large margin nearest neighbor classification,* editor Y. Weiss, B. Schlkopf and J. C. Platt, Advances in neural information processing systems, 18, pp. 1473-1480. MIT Press.

Wolpert, D. H. (1996) *The lack of a priori distinctions between learning algorithms,* Neural Computation, MIT Press, Cambridge, Massachusetts, USA, 8(7), pp. 1341-1390.

Ye, J., Janardan, R. & Li, Q. (2005) *Two-dimensional Linear Discriminant Analysis,* Advances in neural information processing systems, MIT Press, Cambridge, Massachusetts, USA, 17, pp. 1569-1576.

Zhou, Z.-H. (2012) *Ensemble Methods Foundations and Algorithms,* Chapman and Hall / CRC, Boca Raton, Florida, USA, pp. 1-95.