

Increasing Performance of Supervised Machine Learning Methods by Analysis of Construction and Demolition Waste

Petr Kuritcyn, Katharina Anding, Gunther Notni

*Ilmenau University of Technology, Gustav-Kirchhoff-Platz 2, D-98693 Ilmenau, department for quality assurance and industrial image processing,
petr.kuritcyn@tu-ilmenau.de, katharina.anding@tu-ilmenau.de, gunther.notni@tu-ilmenau.de*

Abstract – Any recognition task, where the classes are given by quality rules or standards, needs the use of supervised machine learning. This paper discusses the ways of improvement the performance of methods of spectral analysis and supervised machine learning by classifying the construction and demolition waste (CDW). The first investigations in visible (VIS) and infrared (IR) spectrum have shown, that we can achieve a high recognition rate (98.3%). Therefore, investigations were done for analysing, which methods are useful for improvement classification performance of C&D aggregates.

I. INTRODUCTION

Construction industry is one of the most valued sector of economy in many countries and as many other industries it produces a big amount of waste. In Germany, the construction and demolition waste (CDW) is one of the biggest waste flow with the amount of around 53 million tons per year. The majority of CDW is used in road building and groundwork, while only a small part is recycled into construction materials because of the high heterogeneity of CDW and the mineral admixtures [1]. The application of automatic sensor sorting can solve this problem. This approach is successful in the areas of mining and recycling in Europe and it will become more important in following years, due to increase of waste flow.

The aim is to develop an optical solution for determination of building material classes in CDW by using methods of image processing, spectral analysis and machine learning.

II. STATE OF THE ART

At this time, there existing only devices and methods for selected materials of minerals, ceramics, CDW and waste, for example sorting of calcite and magnesite, limestone and quartz, different crude salts, gravel with different iron content, etc. The given devices using optical and magnetic sensors, NIR- und x-ray sensors in sorting systems with the main focus on high output and not on precision of the analysis. Examples of optical sorting devices for selected

materials can be found for example in [2], [3], [4], [5].

The focus of our research is to develop a CDW analysis method with an optimised supervised machine learning routine to getting high recognition rates for all CDW classes. Our first investigation in space-resolved colour images (see [6], [7]) has shown that we have to use other added spectral information to solve the recognition task in a satisfactory manner. Further investigation in VIS and IR spectrum [8] have shown that high recognition rates (more than 98%) can be achieved by using only spectral information.

Despite this results, we need to improve the recognition rate of some classes to solve this problem in satisfactory manner. Therefore, some investigations of different classifiers, feature selection and extraction methods were done.

III. EXPERIMENTAL BASE OF EVALUATION

The investigations were done on new, not used building materials. They were crushed to obtain homogenous samples. The particle size varies from 8 to 16 mm. Reflectance spectrums were recorded by the IR spectrometer Polytec PSS 2120 (InGaAs detector and spectral range from 1100 to 2100 nm). Our dataset consists of around 1100 spectrums for 9 material classes (lightweight concrete, concrete, aerated concrete, sand-lime brick, dense and porous brick, gypsum, asphalt and granite) with 500 wavelengths (step size – 2 nm).

Supervised machine learning classifiers were used in our classification routine. We have chosen two different classifiers – Random Forest and Support Vector Machine (with polynomial kernel) for our investigations. We used an implementation from machine learning library “Caret” for programming language R [9]. Both classifiers have shown proved stability and high performance in our previous investigations [7, 8] and in many other classification tasks [10] too. Furthermore these two classifiers are well studied [11, 12] and have solid implementations in different programming languages [13, 14, 15]. Besides, Random Forest has few parameters to tune: the number of trees and the number of random

features in tree. The number of trees has to be set large (500 in our case) to reach a plateau with a minimal classification error [11]. The number of random features was set to 2 because of necessity to work with the small number of features in some cases (for example, by the small number of principal components). Support Vector Machines have implementations based on different kernels (linear, gaussian, polynomial etc.). Polynomial kernel is more computationally intensive but delivered more stable results in our investigations without necessity of changing chosen parameters. For polynomial kernel were chosen the following parameters: cost – 1, degree – 2, scale – 0.1. It is important to notice that chosen parameters staying constant during our investigations. It was done to investigate versatility and synergy of classifiers and feature selection methods.

There are plenty methods for feature selection/extraction. We have chosen well studied and less depended on dataset feature selection methods from diversity of methods: with transformation of feature space – Principle Component Analysis (PCA) [16] and Linear Discriminant Analysis (LDA) [17, 18, 19], without transformation of feature space (filter methods) – Information Gain and Chi-Square ranking [20, 21, 22]. One wrapper-method, Correlation Feature Subset Selection (CFS), which is independent of classifier, was also chosen for investigations. In contrast to Information Gain (entropy) and Chi-Square ranking (probabilistic distance), this method is using correlation as evaluation metric [20].

We split our dataset in two parts – training set (60% of data) and test set (rest 40%) for a correct evaluation of classifier performances and to prevent bias. Feature selection methods were applied first on the training set and then, without recomputation, on the test set. The classifiers were trained on the training set using 10-fold cross-validation. After trainings phase, they were tested on the test set.

IV. EVALUATION AND RESULTS

This section is focussed on the comparison of different feature selection/extraction methods in matters of classifier-specific suitability.

The performance of Random Forest and Support Vector Machine (SVM) classifiers were investigated on the pre-process with different feature selection methods data (see Fig. 1).

Firstly, the performance of classifiers was evaluated on the dataset with full feature subset. Random forest classifier achieved the total recognition rate of 87.1%, SVM – 97.4%. This difference may be explained by the fact that SVM kernel transforms original feature space, where classes can be linear inseparable, into high dimensional space, where classes can be linear separable. Random forest lacks in this feature.

Filter selection methods like Chi Square Ranking and Information Gain show no or little improvement in comparison to using full feature subset, but they can be used for reducing the number of features, which leads to reducing the number of recorded wavelengths and requirements to spectrometer. The number of features can be reduced from 500 to 160 without any loss of performance. The total recognition rate of 97.4% was achieved by using SVM classifier on feature subset after applying InfoGain-Filter and 97.9% after applying Chi-Square Ranking filter. Random forest has shown a reduction of the recognition rate of 0.5 – 0.7% by reducing the number of features from 500 to 160 by filter feature selection methods. Further reducing the number of features to 20 leads to significant decreasing of the recognition rate by using SVM classifier -14.1%. The loss by using Random forest classifier is more than twice smaller (4.5 – 6%) in comparison to performance on the full feature subset.

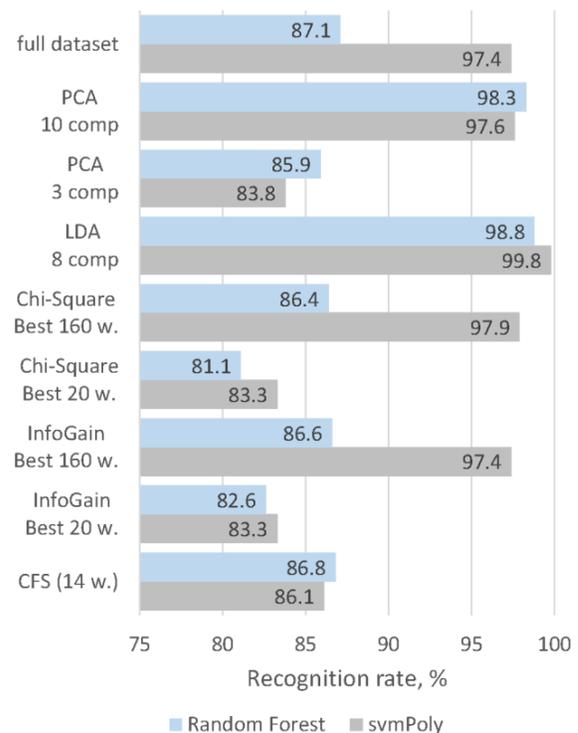


Fig. 1. Comparison of different feature selection methods

Another possibility to reduce feature subset without transformation of feature space is using wrapper methods. Correlation feature subset selection reduced our feature subset to 14 features. Random forest shows a little degradation of performance by using this set – 0.3% (total recognition rate of 86.8%) in comparison to full feature subset. SVM classifier has a bigger drop of performance – 13.3% and a total recognition rate of 86.1%. It has also

another drawback – the complexity of model is raising (340 support vectors against 160 for full feature set).

Methods with a transformation of feature space show an increase of total recognition rates for both classifiers. We used principal component analysis as pre-processing step in our previous investigation in combination with random forest classifier and it has shown the high recognition rate of 98.3% by applying on feature subset, which consists of 10 principle components [8]. SVM classifier shows small improvement in the case of 10 principle components in comparison to applying on the full feature subset – 97.6% against 97.4%. PCA provides a feature space transformation, which SVM already has. In our case, it does not make the classification model easier: SVM model for a full feature subset consists of around 160 support vectors and the model for a feature subset of 10 principle components - around 180 support vectors. But PCA as a pre-processing step has a big impact on performance of random forest classifier: it improves the total recognition rate by 12% and makes random forest

competitive against SVM. Further reducing the number of principal components leads to degradation of classifiers performance. The first three principle components provide too little information for a satisfactory performance of classifiers. Random forest achieved the total recognition rate of 85.9% and SVM of 83.8%. It is worth to notice that SVM shows poor performance on small feature subsets – low dimensionality of data eliminates advantages of SVM with kernel (transformation of feature space).

An alternative for PCA is Linear Discriminant Analysis, which, on the contrary, uses class-specific information. While PCA tries to retain most of variability in dataset, LDA tries to retain most of interclass variance that can lead to an improvement of classification performance by using less number of features. Applying LDA on our dataset produced 8 linear discriminants (9 classes minus 1). Fig. 2 shows results of this analysis in form of projection on different linear discriminants (the percentage in brackets is an information value of current linear discriminant).

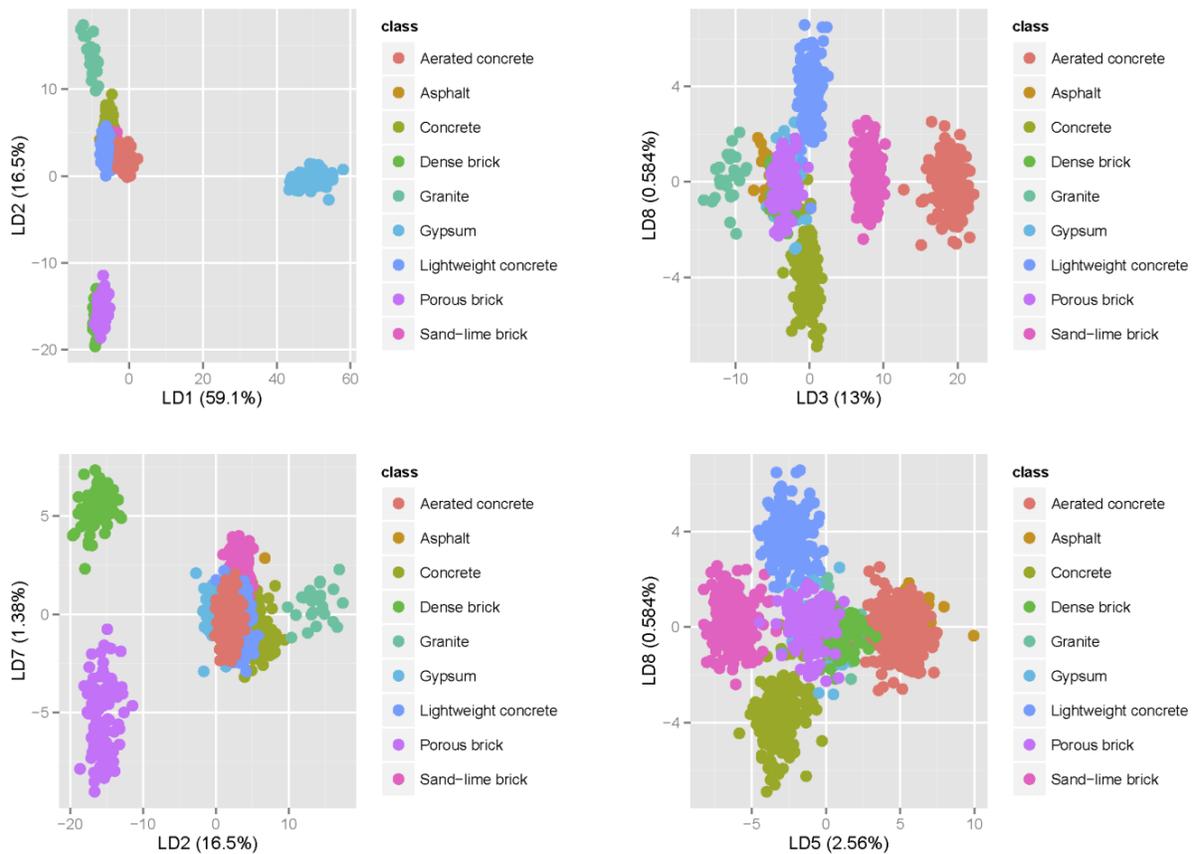


Fig. 2. Visualization of results of linear discriminant analysis on NIR-spectrum dataset of CDW

Despite the fact that class-specific clusters are overlapped by the first three linear discriminants like by using the first three principle components [7], there existing combinations of linear discriminants, which show a good separation of all classes in our dataset. The first and second linear discriminants show the following autonomous clusters – granite and gypsum; second and seventh - dense and porous brick; third and eighth – sand-lime brick and aerated concrete; fifth and eighth – concrete and lightweight concrete.

Using LDA as a pre-processing step shows a good synergy with both classifiers and results in the total recognition rate of 98.8% by using Random Forest as classifier. SVM classifier achieved even better results with 99.8%.

V. CONCLUSIONS

The best results were achieved by using Linear Discriminant Analysis as a pre-processing step. Classifiers Random Forest and SVM with polynomial kernel show high performance in combination with this feature selection method of 98.8% and 99.8%. These results exceeded the previous best results with PCA (98.3% by using Random Forest). One disadvantage of this approach is the necessity of recording complete spectrums with full resolution, i.e., all 500 wavelengths.

Filter selection methods like Chi Square Ranking or Information Gain can be used for reducing the number of features from 500 to 160 in subset without any loss of performance in comparison to the full feature subset. It leads to decreasing the number of recorded wavelengths and, as a result, it decreases requirements to spectrometers/cameras.

The wrapper method, Correlation Feature Subset Selection, shows a weaker performance and should not be used for further investigations on spectrums of CDW.

Classifiers due to their nature have shown different behaviour in our investigations:

- Random Forest classifier has shown good synergy effects with feature selection methods, which transform original feature space (principal component analysis, linear discriminant analysis).
- SVM with polynomial kernel has shown best synergy effects with filter selection methods. It can be used in scenario with a detector, which consists a relatively small array of cells. Also SVM has shown exceptional performance in combination with Linear Discriminant Analysis.

VI. ACKNOWLEDGEMENT

The basis of the captured image and spectral datasets builds a big amount of CDW samples, sorted by the Dr.-Ing. Elske Linß from Faculty Civil Engineering of Bauhaus-University of Weimar, Germany. We thank for the material support.

The PhD grant for Petr Kuritcyn, which forms the basis

of this paper, is funded by German federal state of Thuringia. We thank for the financial support.

REFERENCES

- [1] „Mineralische Bauabfälle Monitoring 2010“, Arbeitsgemeinschaft Kreislaufwirtschaft Bau (KWB), Berlin, 2013, pp. 8–12.
- [2] M. Dehler, “Optische Sortierung von mineralischen Rohstoffen“, *Schüttgut*, Vol.: 10 (3), Seiten: S.212-216, 2004.
- [3] H. Harbeck, “Processing of Aluminum Scrap with sensor-based sorting, *Aufbereitungstechnik*, Vol.: 47 (12), Seiten: S. 4-12, 2006.
- [4] P. J. Groot et al., “Validation of remote, on-line, near-infrared measurements for the classification of demolition waste”, *Analytica Chimica Acta*, Vol.: 453, Seiten: p.117–124, 2002.
- [5] C. F. Hendriks et al., “Suitable Separation Treatment of Stony Components in Construction and Demolition Waste (CDW)”, *International RILEM Conference on the Use of Recycled Materials in Building and Structures*, 8. – 11. November 2004, Barcelona, Volume 1, S. 166 – 172.
- [6] K. Anding, E. Linß, H. Träger, M. Rückwardt, and A. Göpfert, “Optical identification of construction and demolition waste by using image processing and machine learning methods”, *14th Joint Int. IMEKO Symp.*, Jena, Germany, 2011.
- [7] P. Kuritcyn, K. Anding, E. Linß, S.M. Latyev, “Increasing the safety in recycling of construction and demolition waste by using supervised machine learning”, *IMEKO Joint Symposium 2014 TC1 - TC7 - TC13*, Madeira, Funchal, September 2014.
- [8] K. Anding, P. Kuritcyn, E. Linß, S.M. Latyev, “Significant Characteristics in VIS- and IR-Spectrum of Construction and Demolition Waste for High-Precision Supervised Classification”, *OCM 2015 - Optical Characterization of Materials - conference proceedings*, KIT Scientific Publishing, Karlsruhe 2015.
- [9] M. Kuhn. Contributions from J. Wing, S. Weston, A. Williams, C. Keefer and A. Engelhardt (2012), “caret: Classification and Regression Training”, R package version 5.15-044. <http://CRAN.R-project.org/package=caret>.
- [10] M. Fernández-Delgado, E. Cernadas, S. Barro, D. Amorim, “Do we need hundreds of classifiers to solve real world classification problems?”, *J. Mach. Learn. Res.* 15, 1 (January 2014), 3133-3181.
- [11] L. Breiman, “Random Forests”, *Mach. Learn.* 45, 1, October 2001, 5-32.
- [12] C. Cortes, V. Vapnik, "Support-Vector Networks", *Machine Learning*, 20, pp.273-297, 1995.
- [13] M. Hall, E. Frank, G. Holmes, B. Pfahringer, P. Reutemann, I.H. Witten, “The WEKA Data Mining Software: An Update”, *SIGKDD Explorations*,

- Volume 11, Issue 1, 2009.
- [14] C.-C. Chang, C.-J. Lin, “LIBSVM: a library for support vector machines”, *ACM Transactions on Intelligent Systems and Technology*, 2:27:1--27:27, 2011.
 - [15] Pedregosa et al., “Scikit-learn: Machine Learning in Python”, *JMLR* 12, pp. 2825-2830, 2011.
 - [16] D.A. Burns, E.W. Ciurczak, “Handbook of Near-Infrared Analysis”, Third Edition, CRC Press, 2007
 - [17] X. Shao, W. Zheng, Z. Huang, “In vivo diagnosis of colonic precancer and cancer using near-infrared autofluorescence spectroscopy and biochemical modelling”, *J. Biomed. Opt.*; 16(6), 2011.
 - [18] R.O. Duda, P.E. Hart, D.G. Stork, “Pattern Classification”, 2nd Edition. Wiley-Interscience, 2000.
 - [19] C.M. Bishop, “Pattern Recognition and Machine Learning”, Information Science and Statistics, Springer-Verlag New York, Inc., Secaucus, NJ, USA, 2006.
 - [20] I. Guyon, S. Gunn, M. Nikravesh, .L.A. Zadeh, “Feature Extraction: Foundations and Applications”, Studies in Fuzziness and Soft Computing, Springer-Verlag New York, Inc., Secaucus, NJ, USA, 2006.
 - [21] C.C. Aggarwal, “Data Mining: The Textbook”, Springer Publishing Company, Incorporated, 2015.
 - [22] C.C. Aggarwal, “Data Classification: Algorithms and Applications”, 1st ed., Chapman & Hall/CRC, 2014.