Analysis of Ensemble methods applied to Lithology Classification from Well Logs

Vanessa R. C. Leite*, Pedro Mário C. Silva, Marcelo Gattass, Tecgraf/PUC-Rio, Brazil, Aristófanes C. Silva, UFMA, Brazil

This paper was prepared for presentation at the 13th International Congress of the Brazilian Geophysical Society, held in Rio de Janeiro, Brazil, August 26-29, 2013.

Abstract

Lithology classification is an important task in reservoir characterization, one of its major purposes is to support well planning and drilling activities. Therefore, faster and more effective classification algorithms will increase the speed and reliability of decisions made by geologists and geophysicists. This paper analyzes ensemble methods applied to automatic lithology classification. For this, we performed a comparison between single classifiers (Support Vector Machine and Multilayer Perceptron) and these classifiers with ensemble methods (Bagging and Boost). The results are very satisfactory, and confirm the advantages of using ensemble methods. However the trade-off between performance improvements versus resource utilization shows that the use of ensemble methods is only necessary when precision is an extremely determinant factor.

Introduction

Lithology classification is commonly performed by two ways (Thomas, 2004): core analysis and manual analysis of well logs by an experienced geologist. In the core analysis classification, the lithology is identified during the drilling process and generates accurate results. However, it is an expensive process and consequently impossible to be performed at almost all wells. On the other hand, most drilled wells have available well logs, which may be used for classification, without the need of core information. Well logs are obtained through physical measurements (electrical resistivity, electric potential, natural or induced radioactivity, among others) made by instruments lowed into the hole. Figure **??** shows two well logs and the respective lithology. Lithology classification from well logs is usually a manual process. Many works such as (Al-Anazi et al., 2010), (Hsieh et al., 2005), (Santos et al, 2002), (Santos et al, 2003), among others, aim to automate the lithology classification process. This automation is generally made using machine learning techniques. Such techniques receive as input a training data set that has been classified manually. Thus, with complete information (input data set already identified), these technique try to learn how to classify the same data and also how to classify new entries.

Al-Anazi et al. (2010) used three classifiers: Linear Discriminant Analysis (LDA), Support Vector Machine (SVM) and Probabilistic Neural network (PNN) in other to compare the obtained results. To classify three lithology types, they used four well logs: gamma ray (GR), caliper log (CAL), neutron porosity (NPHI), and photoelectric log (PEF). The three classifiers obtained similar rates and SVM results were superior to 95%. In (Hsieh et. al, 2005) is made lithology identification of aquifers from geophysical well logs and fuzzy logic analysis. To construct a fuzzy lithology system they used four well logs: gamma ray, borehole compensated sonic (BHC) with sonic porosity (SPHI) curve, spontaneous potential, and phasor induction (PI), which is the most useful log in this kind of study. The identification is based on an inference system composed by rules, which determine the probability of a set of entries (well logs) belongs to a lithology type. The authors identified five lithology types: silt, clay, fine sand, medium sand e coarse sand. The results were fairly good, reaching a rate of 90%. Recently, ensemble techniques have also been used in order to improve the results of lithology classification based on machine learning techniques. In (Santos et al., 2002) and (Santos et al., 2003) are used ensemble methods such as Driven Pattern Replication (DPR) and ARC-X4 with MLP, to lithology classification of stratified data and non-stratified data. They used four well logs: gamma ray, sonic, density and resistivity and plus the observation's depth, totalizing five attributes. The authors identify in the first paper eight distinct lithologies, and in the second three categories of rock types. It is demonstrated that in both types of data ensemble

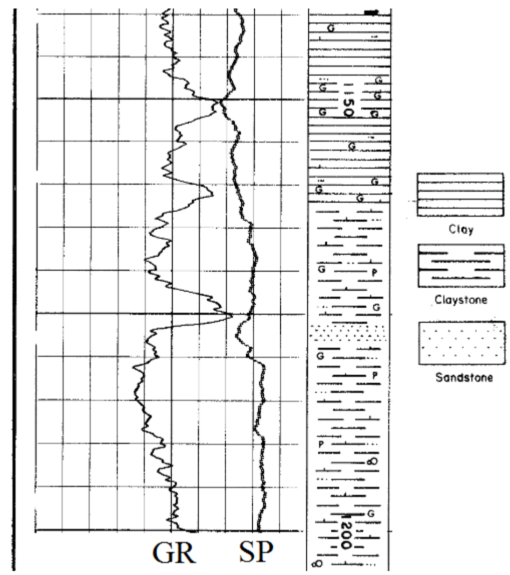

Figure 1: Well logs examples: Gama Ray (GR) e Spontaneous Potential (SP), and their respective lithology. The legend of lithology type is in the right side of figure.

methods improve the results of the single classifiers. Following a similar approach, the goal of this paper is to compare lithology classification by machine learning, using the single classifiers most promising such as Support Vector Machine (SVM) and Multilayer Perceptron (MLP), versus the popular ensemble methods, specially, Bagging e Boost. Moreover, we aim to distinguish seventeen lithologies, which are divided into different categories such as sediments, sedimentary rocks, and salt, among others.

Methodology

To perform the lithology classification and identify the importance of classification by ensemble methods, several experiments were conducted. We used two different classifiers: MLP and SVM, and two different ensemble methods: Bagging and Adaboost. SVM is a machine learning technique proposed by Vapnik (1995) that uses statistical learning. SVM constructs a hyperplane used to separate sets. The best result is achieved by the hyperplane that has the largest distance to the nearest training data point of any set. SVM was originally developed as a method of linear separation (Cortes and Vapnik, 1995). It is possible to extend SVM to separate sets that are non-linear carrying the data to a higher dimensional space in which they may be separated linearly. MLP is an artificial neural network (ANN). ANN is composed of multiple layers of nodes in a directed graph, with each layer fully connected to the next one. Note that there is no edge connecting the input layer to output layer. MLP consists of three or more layers: one input layer, one or more hidden layers and one output layer. Generally, data is presented at the input layer, the network then process the input in the hidden layer. The output layer provides the final result, i.e., that class the best to fit to the input data. Moreover, MLP is a feedforward neural network, that is, data flows in one direction from input to output (Noriega, 2005). Ensemble methods represent a set of classifiers whose individual decisions are combined in some way to classify new examples. By combining solutions we can improve the results because the reduction in error can be viewed as arising from reduced variance due to the averaging over many solutions (Dietterich, 2000). In this work we use the Adaboost (Freund and Schapire, 1996) and the Bagging (Optiz et al., 1999) algorithms. There are many ways to construct ensemble, such as: enumerating the hypotheses, manipulating the training examples, manipulating input features, among others. In this paper, both algorithms used are based on training example manipulation (Dietterich, 2000). Adaboost algorithm is based on the fact that the performance of single classifiers is improved when they are iteratively combined, that is, each subsequent classifier will pay more attention to the samples incorrectly classified by the earlier classifier. This way, given a database, each sample is considered to have the same initial weight $(1/number\ of\ samples)$. The classifier is then built for that database and in the end for each sample incorrectly classified, the weight is summed, defining ε. Each misclassified sample has its weight updated by a multiplier factor defined by $(1-\varepsilon)/\varepsilon$. After updating these samples, the rest of the samples is updated in a way that the sum of all weights is equal to one. Once again the classifier is built and the process is repeated until the number of repetitions previously defined is reached. The Bagging algorithm is based on the fact that great part of classifying errors is due to very specific training database choice. This way, the basic idea is to generate various training sets from the original database. Each new set may contain an already included sample or may not contain some samples. Therefore, for each set created a new function capable of classifying that set is generated. So, given a DB database of size N, several training sets DBt are created. Each training set can have a different number of elements. Then, a Ct classifier is created for each DBt and trained. So, for classifying a sample, each Ct classifier returns its prediction that counts as one vote. The final classifier Cf counts every vote and determines the answer as the class with most of the votes. This way, we present a lithology classifier through single classifiers, SVM and MLP, and lithology classifier using the addiction of ensemble algorithms, Adaboost and Bagging, in a way that we have six experiments for lithology classification: single SVM, single MLP, SVM plus Adaboost, SVM plus Bagging, MLP plus Adaboost, MLP plus Bagging. More information about the tests realized is discussed in the next section. To perform the experiments a machine learning simulation tool developed by Waikato University called Waikato Environment for Knowledge Analysis - Weka (Hall, M et al., 2009) was used. This tool contains a collection of machine learning algorithms licensed under GPL (General Public License). The SVM algorithm used in this tool is LibSVM (Chang and Lin, 2011). A cross-validation technique was also used, as this method generates more reliable data than a simple division between test set and training set. Cross-validation makes possible to use all the data for training and tests in alternating ways, since the set is divided into groups. A group is set apart for classification and the others are used for training. This process is repeated for each group and the accumulated rate error is calculated (Hansen and Salamon, 1990). In this work was used 10-fold cross-validation. To evaluate the results confusion matrix, precision, recall and f-measure measurements were used (Fawcett, 2005), (Rijsbergen, 1979).

Experiments and Results

For this work, we used well log information from the public database of the North Sea, Geological Survey of the Netherlands (available at: `http://www.nlog.nl/nlog/listAllWellLocations`). The wells utilized are part of the F set, which contains 108 wells, all of them with latitude and longitude information. The wells were mapped to their locations for a well-distributed spatial selection because wells from the same area are likely to have similar characteristics, which would not generate a well-diversified base. From the 108 wells, 11 were selected for acquisition of samples of well logs and lithology. The needed information (well logs and lithology) were separated, well logs in a text file and lithology in an image file. The conversion of lithology information to a text file was manually done. Each sample contains information on well depth and five electric profiles: gamma ray (GR), neutron porosity (NPHI), sonic log (DT), density log (RHOB) and compensation density curve log (DRHO). The converted data is available at: `http://www.tecgraf.puc-rio.br/welllogs/data/ArffFiles.zip`. After converting the data, two groups were created, both divided in 17 classes (i.e., lithologies): Group 1 with 53.181 samples;

Group 2 with 9.996 samples. The second group is a subset of the first, created for better assessing the behavior of the Adaboost algorithm, as its principal characteristic is to replicate the database. The division of the samples by lithology type is presented at Table **??**.

Table 1: Lithologic types and number of samples for each group.

Lithologic type	Number of samples (Group 1)	Number of samples (Group 2)
mudstone	5679	795
anhydrite	830	677
claystone	7877	896
limestone	3713	201
chalk	5963	179
sandstone	2861	736
sand	407	407
argilaceous	1670	536
calcareous	795	795
chert	74	74
shale	1554	522
tuff	344	344
clay	13681	1165
shale/sandstone	2822	659
marl	904	407
sandy	3687	536
silty	320	320

The tests were performed using the following parameters: in LibSVM the kernel function used was the radial basis function, with parameters c = 8 e γ = 8192 for group 1 and c = 8 and γ = 512 for group 2. In the MLP, a hidden layer with 11 nodes was used and the learning rate was 0.3, and the momentum was 0.2 Several tests were done to analyses the classifiers, both individually and using ensemble methods. Thus, twelve experiments were realized: six with Group 1 and six with Group 2. The experiments were executed in a 64-bits architecture with 4GB RAM, Intel Core 2 Duo 2.40GHz processor. Based on confusion matrix, performance measurements can be calculated. In addition, this matrix is able to show how the system confuses the resulting classes. The confusion matrices obtained on this work are available at: http://www.inf.puc-rio.br/ vleite/confusionMatrices.html. The time, in minutes, for each of the twelve experiments is available in the Table **??**.

Table 2: Time, in minutes, obtained on realization of the experiments.

Samples		Single	Adaboost	Bagging
Group 1	SVM	272	15148	2870
	MLP	89	598	707
Group 2	SVM	1	254	194
	MLP	12	84	138

The amount of correctly classified instances, i.e. the recall rate, in the experiments is available in Table **??**. The results of f-measure and precision rates are available in Table **??** and Table **??**.

Table 3: Time, in minutes, obtained on realization of the experiments.

Samples		Single	Adaboost	Bagging
Group 1	SVM	90.21%	90.78%	90.25%
	MLP	72.76%	72.76%	75.68%
Group 2	SVM	93.39%	94.78%	93.33%
	MLP	69.26%	77.73%	73.54%

Table 4: Time, in minutes, obtained on realization of the experiments.

Samples		Single	Adaboost	Bagging
Group 1	SVM	90.1%	90.7%	90.1%
	MLP	71.1%	71.1%	73.6%
Group 2	SVM	93.3%	94.8%	93.2%
	MLP	69.2%	77.5%	73.5%

Table 5: Time, in minutes, obtained on realization of the experiments.

Samples		Single	Adaboost	Bagging
Group 1	SVM	90.%	90.8%	90.3%
	MLP	71.8%	71.8%	75.1%
Group 2	SVM	93.5%	94.8%	93.4%
	MLP	70.1%	77.7%	74.8%

Analyzing the results, we firstly noticed that the SVM results were always better than the MLP results, even if we compare the performance measurements of single SVM classifier versus MLP using ensemble methods. This variation denotes that the SVM is better at the classifying task. One of the factors that can explain the result is that the MLP implements a global approximation strategy, while SVM uses a local approximation, besides having a statistical learning formulation. Based on the rates presented, it can be noticed that the recall rate for the MLP was better in the group with more information for training, although when ensemble methods were used, there was an improvement in the results, with exception of group 1 with Adaboost. This suggests that MLP networks do not improve when many data with similar information are used, although in group 2, the improvements made by Adaboost are around 7%. The SVM is a better classifier for lithology identification even when it has large lithological variation, and the use of ensemble methods provided small improvements in the performance rates. In some cases, it obtained the same rates of the single classifier, making the resource consumption the greater difference, which makes the use of ensemble methods with SVM practically impracticable. When we observe the confusion matrix, we note that the small improvement generated by the ensemble methods for SVM is due to the fact that some classes are hard to distinguish. These same classes are also confused by the MLP. By analyzing, we note that the confused classes are, mostly, sediments that will make up sedimentary rocks.

When comparing ensemble methods, we noticed that

Adaboost had better performance than Bagging. We can attribute this improvement to the way that they are built. As the Adaboost generates new samples for classifying, always pay more attention in samples that were incorrectly classified, in specific cases, it turns out to be better, while the Bagging method creates subsets of the database, which may contain an already included sample or may not contain some samples. These characteristics make Adaboost more sensitive to data noise, so in a database with high noise rate, the Bagging method should present better results.

Conclusions

According to the results of the experiments, we noticed that a simple SVM classifier provides better results than a MLP neural network. However, both ensemble methods (Bagging and Adaboost) provide improvements in the lithology classification results when utilizing a MLP neural network. This improvement reached rates between 4% and 7%, which can justify the choice of ensemble methods in spite of the greater resources cost for obtaining results. The use of SVM demonstrated results much more satisfactory than the use of MLP. SVM obtained low improvements when comparing the results of the single classifier against the use of ensemble methods. Therefore, the trade-off between precision and the resources consumption shows, in this initial study, that the use of ensemble methods for support vector machines is not worthwhile. As future works, we propose to analyze a noisier database to assess the robustness of the Bagging method. We also want to improve the differentiation between the classes that represents sediments.

References

Al-Anazi, A. and I. Gates, 2010, On the Capability of Support Vector Machines to classify lithology from well logs: Natural Resources Research, 19, no. 2, 125-139.

Chang, C.-C. and C.-J. Lin, 2011, LIBSVM: a library for support vector machines. ACM Transactions on Intelligent Systems and Technology, 2:27:1–27:27.

Cortes, C. and V. Vapnik, 1995, Support-vector network. Machine Learning, 20, 273-297.

Dietterich, T. G., 2000, Ensemble methods in machine learning, in J. Kittler and F. Roli, eds., Multiple Classifier Systems 1857, 1-15.

Fawcett, T., 2005, An introduction to ROC analysis: Pattern Recognition Letter, 27, 861-874.

Freund, Y. and R. E. Schapire, 1996, Experiments with a new boosting algorithm: Machine Learning Proceedings of the Thirteenth International Conference, 148-156.

Hall, M., E. Frank, G. Holmes, B. Pfahringer, P. Reutemann, and I. H. Witten, 2009, The WEKA Data Mining Software: An Update, SIGKDD Explorations.

Hansen, L. and P. Salamon, 1990, Neural network ensembles: IEEE Transactions on Pattern Analysis and Machine Intelligence, 12, no. 10, 993-1001.

Rijsbergen, C. J. van, 1979, Information Retrieval: Butterworth-Heinemann

Santos R. O. V. dos, F. A. V. Artola, M. M. B. R. Vellasco and S. A. B. Fountoura, 2002, Lithology recognition by neural networks, XVI Brazilian Symposium on Artificial Intelligence - Lecture Notes in Artificial Intelligence

Santos R. O. V. dos, M. M. B. R. Vellasco, F. A. V. Artola and S. A. B. Fountoura, 2003, Neural net ensembles for lithology recognition, IV Multiclassifier System Conference, LCNS 2709, 246-255

Noriega L., 2005, Multilayer Perceptron tutorial: School of Computing, Staordshire University

D.Optiz and R. Maxlin, 1999, Popular ensemble methods: an empirical study: Journal of Artificial Intelligence Research, 11, 169-198.

Thomas, J. E., 2004, Fundamentos de engenharia do petróleo: Interciência.

Vapnik, V., 1995, The nature of statistical learning theory: Springer.

Acknowledgments

The authors would like to thank the v3o2 software development team for the ideas and help in conversion of lithological image files. Vanessa Leite would like to thank CAPES (Coordenação de Aperfeiçoamento de Pessoal de Nível Superior) for scholarship support.