Computational Intelligence Study of the Iron Age Glass Data

Karol Grudziński
Department of Physics
Bydgoszcz Academy
Plac Weyssenhoffa 11
85-072 Bydgoszcz, Poland
kagru@phys.uni.torun.pl

Maciej Karwowski
Institute of Archaeology
University of Rzeszow
36-007 Krasne 32a, Poland
mkar@univ.rzeszow.pl

Włodzisław Duch
School of Computer Engineering
Nanyang Technological University, Singapore,
and Department of Informatics,
Nicholaus Copernicus University, Toruń
Poland, duch@ieee.org

Abstract— **Relative abundance of chemical elements allows for classification of the Late Iron Age glass samples to one of the three main chronological periods (LT C1, LT C2 and LT D1) of glass artifacts. Predictive abilities of various classification systems, including several rule-discovery methods, have been compared in this paper. The results indicate the usefulness of machine-learning methods in such applications. A hypothesis stating that the glass surface corrosion has minor influence on the results of chemical analysis has been confirmed by computational intelligence methods.**

Keywords: classification, machine learning, data-mining, archeological glass, dating of glass artifacts.

I. INTRODUCTION.

Analysis of the archeological glass data has many aspects. This paper is concerned with estimation of the chronology of the artifacts studied, using spectroscopic analysis of the glass composition. Possibility of dating artifacts using spectroscopic analysis would be of great importance for archeology. Concentration of many chemical compounds may be measured using such methods, but only some of these components have high influence on prediction of the age of glass artifacts. From the machine learning perspective this means that rule-based methods with selection of attributes will be more informative than predictive statistical methods. This data presents an interesting challenge, since it contains several samples from the same object, confusing some classification methods that relay on similarity of samples.

The glass composition has been measured usually in several places: on the original surface of the artifact and on the broken parts. Therefore in the original database several instances correspond to a single glass object. A thin corrosion layer always covers the surface of the archeological glass, and the broken parts are usually cleaner or at least less corroded. It is however not clear how much the surface corrosion influences the results of the measurements. This are interesting questions that, at least in respect to artifacts dating, could be answered using a data-mining approach.

In the next section a description of the database is given. In the third section the methodology underlying this study is presented, and the results of a number of classifiers compared. Last section concludes the paper.

II. THE DESCRIPTION OF THE DATABASE

The archeological glass database has been obtained during realization of an interdisciplinary project "Celtic Glass Characterization", under supervision of Prof. G. Trnka (Institute of Prehistory, University of Vienna) and Prof. P. Wobrauschek (Atomic Institute of the Austrian Universities in Vienna). The measurements of chemical compound concentrations were made using the Energy Dispersive X-ray Fluorescence Spectroscopy [1], [2]. Concentration of the following 26 compounds have been measured: Na_2O, MgO, Al_2O_3, SiO_2, SO_3, K_2O, CaO, TiO_2, Cr_2O_3, MnO, Fe_2O_3, CoO, NiO, CuO, ZnO, SeO_3, Br_2O_7, Rb_2O, SrO, ZrO_2 MoO_3, CdO, $SnO2$, Sb_2O_3, BaO, and PbO.

The original database consists of the description of 555 glass samples. Those of them which are from unknown time period, correspond to glass artifacts of uncertain chronology, or archeologically do not belong to the rest of the data, have been excluded from the database (see detailed description in the next section). In order to prepare datasets for our computational studies we have excluded also from the original dataset relatively small number of samples containing measurements of additional decorations on glass. Those decorations are usually made of a different kind of glass than the main body of the glass artifact. Since chemical analysis has been made for most glass samples in several areas of the glass, several entries in the database may correspond to a single glass object. Usually two measurements are made on the surface and one on each of the two broken sides.

Three main chronological periods are of interest to archeologists:

1) LT C1 - La Tene C1 period, 260 - 170 B.C.
2) LT C2 - La Tene C2 period, 170 - 110 B.C.
3) LT D1 - La Tene D1 period, 110 - 50 B.C.

For the prediction of the chronology of studied objects, experiments on various subsets of the original data have been performed. First experiment has been performed on data containing both surface and broken part measurements. The second and the third experiment has been performed separately for data consisting of surface and of broken parts measurements. This has been done to check if the place

of measurement is significant for chemical dating of glass artifacts.

Experiments conducted with similarity based classification methods using cross-validation tests give very high accuracies. This is a general problem for data that contains repeated samples from the same objects: since the most similar sample comes from the same object, classification using the nearest neighbor is perfect. This problem cannot be solved using standard statistical techniques, such as bootstrap evaluation of accuracy [7] instead of crossvalidation, since the probability of finding samples from the same object in the training and the test partitions will be high (and difficult to evaluate). The only reasonable solution is to use the low-complexity models, such as simple decision trees, logical rules or neural networks with low number of hidden neurons and strong regularization. Such models provide simple decision borders and should not suffer significantly from this problem, although realistic evaluation of the expected accuracy remains difficult due to strong correlation of the samples.

To alleviate the problem of unbiased accuracy estimation each of the datasets has been divided into training and test partition to assure that distinct samples that are almost identical und thus presumably come from the same glass object have not been included in the same training partition. There are approximately 4 samples per glass object. Still classification accuracy of the algorithms trained through cross-validation in case of the first experiment may be too high, because in both the training and the test sets there are usually two samples taken from a side and a surface of the same glass object. In such situation training of the classification model becomes an uncontrolled bootstrap process. In case of the second and third experiment there is only one case belonging to a single glass object but experiments conducted on this data still confuse similarity based models - this time only at the testing phase of the classification process.

We have several computational intelligence tools for this study: the Waikato Environment for Knowledge Analysis (WEKA) [4], the NETLAB software for Matlab environment [3], the Similarity Based Learner system (SBL) [5], and the Ghostminer data mining software [6]. The last two software systems have been developed in our laboratories.

III. NUMERICAL EXPERIMENTS

In each subsection a distinct computational experiment on differently partitioned data is described.

A. The First Experiment (Surface and Broken Side Data)

The first study has been performed on the data containing measurements made on both original surface and the broken parts. The sample distribution among classes for this experiment is:

1) LT C1, 29.68% (84 cases),
2) LT C2, 33.57% (95 cases),
3) LT D1, 36.75% (104 cases).

The total number of cases is 283, with the base rate of 37%. 143 samples have been selected as the training partition and the remaining 140 as test set. Table (I) summarizes the best results obtained for this data.

TABLE I

RESULTS FOR THE SURFACE AND BROKEN SIDE DATA.

System	Train %	Test %
Naive Bayes (WEKA)	81.8	81.4
1-NN (SBL)	100.0	75.0
IncNet, 5 neurons (GM)	99.3	73.6
SSV - Tree (GM)	97.1	70.5± 2.9
MLP+ regulariz. (Netlab)	98.6	70.0
MLP backprop. (Netlab)	100.0	67.1
1R (WEKA)	74.1	66.4
SVM (GM)	99.3	63.6
C4.5 - Rules (WEKA)	91.6	62.1
C4.5 - Tree (WEKA)	97.9	55.7

There are no problems with distinguishing objects from LT C1 period: only 3 test samples from this class were assigned to other classes by Naive Bayes, and all were correctly recognized by MLP with regularization ($\alpha = 0.1$ used has speeded up convergence), achieving 100% sensitivity. The second class has lowest sensitivity (67.4% for Naive Bayes and only 21-31% for MLP), while the third class has moderate 84% sensitivity for Naive Bayes and 88% for MLP. Similar, although slightly more accurate results were obtained using the IncNet incremental RBF-like network [10]. Visual inspection of this dataset was done using the GhostMiner multidimensional scaling software [6]. Fig. 1 shows that a good separation of the LT C1 class from the other two may be expected.

The similarity based methods may be especially misleading, since in the training and test there are two cases belonging to the same glass object. Therefore we have avoided more sophisticated SBL methods, restricting optimization only to metric and k. The cross-validation results on the training partition are much better then on the test set probably because of the uncontrolled bootstrap learning. Results of the SVM are relatively poor although bias and C coefficient has been fully optimized and various kernels tried [9]. The best results are obtained with Naive Bayes classifier, but they may suffer from the same problem as SBL.

1R decision tree gives very simple decision rules, based on selection of a single attribute that allows for the best classification, and defining the intervals where samples from a single class are prevalent. The most informative attribute for the prediction of the chronology of the glass found by the 1R decision tree is the concentration of MnO. The rules using this one attribute are:

IF $MnO < 2185.205$ THEN C1
IF $MnO \in [2185.205, 9317.315)$ THEN C2
IF $MnO \geq 9317.315$ THEN D1

These rules predict correctly 100 out of the 143 training samples, and 93 out of 140 test samples.

Other attributes that were found to be important are concentration of TiO_2, Fe_2O_3, NiO, MnO, Sb_2O_3 and ZnO for glass from LT C1 period, concentration of Fe_2O_3, TiO_2, NiO, and
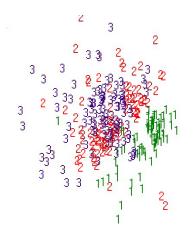
Fig. 1. Glass data from the first experiment plotted in two MDS dimensions. The LT C1 class is quite well separated from LT C2 and LT D1 classes.

System	Training %	Test %
1-NN (GM, norm, Manh)	100.0	94.1
1-NN (GM, normalized, Euclid)	100.0	89.7
MLP (16 neurons, WEKA)	100.0	86.8
IncNet (3 neurons, GM)	100.0	86.8
1-NN (GM, std, Euclid)	100.0	85.3
SVM (GM, Gauss kernel)	98.4	85.3
SSV Tree (GM, opt prune)	100.0	83.8
SSV Tree (GM, opt prune)	90.2	82.4
C4.5 - Rules (WEKA)	96.7	79.4
NaiveBayes + 8 attr sel (WEKA)	86.9	77.9
C4.5 - Tree (WEKA)	95.1	73.5
NaiveBayes (WEKA)	78.7	72.0
1R (WEKA)	72.1	63.2

PbO for glass from the LT C2 period and TiO_2, Sb_2O_3, Fe_2O_3, PbO, and ZnO for glass of the LT D1 period. Simple selection procedure (forward and backward wrapper approaches) for Naive Bayes has identified four attributes, MnO, SrO, ZrO_2 and PbO; two of these attributes do not contribute much, and the results are significantly degraded. Genetic optimization of the attribute set leads adds two more attributes, Sb_2O_3, and Al_2O_3, but the test set results are still about 5% lower than without attribute selection.

MnO has also been identified as the most important attribute by the SSV decision tree [6], [8]. Since the tree is constructed using internal 5-fold crossvalidation on the training set (the number of folds has been optimized to reach the highest accuracy on the training set) it is not deterministic, achieving slightly more than 70%±2.9% on the test. The final trees use only 9 features, have between 30-40 nodes, corresponding to about 15 rules. The complexity increase over the 1R tree is substantial, and the gain in accuracy on the training set large, but the training results are only slightly better. We have also used other models available in the WEKA package, but results were usually significantly worse.

The best model may be selected using crossvalidation calculation on the whole dataset.

B. The Second Experiment (Surface Data)

For the second experiment only samples with measurements on the glass surface were selected. There are 129 such instances in the database. The distribution of classes is:

1) LT C1, 26.36% (34 cases),
2) LT C2, 37.98% (49 cases),
3) LT D1, 35.66% (46 cases).

We divided them into 61-case training partition and took the remaining 68 as the test partition. The splitting was made in such a way that cases obtained from measurements on the same glass object are separated. The table below summarizes the best results obtained for this data.

The single-attribute rules obtained from 1R system on this data are:

IF $MnO < 187.34$ Then C1
IF $MnO \leq 9489.09$ Then C2
IF $MnO \in [3821.99, 9489.09) \vee MnO \geq 9489.09$ Then D1

The bucket size for the 1R algorithm has been was taken as 8, a value optimized using 5-fold crossvalidation on the training set. These rules handle correctly 42 out of 61 training samples and 43 out of 68 test instances. Principal component analysis required 16 components to cover 95% of variance. MLP results with these components, with feedforward wrapper feature selection and without any selection are very similar, correctly classifying 59 out of 68 test cases.

C4.5 decision tree has used 7 attributes, creating 17 nodes, including 9 leaves. C4.5rule version provides independent logical rules that obtain 79.4% accuracy on the test set. These rules are listed below:

If $ZrO_2 > 296.1$ Then C1 (16/0)
If $Na2_O \leq 36472.22$ Then C1 (2/0)
If $Sb2O3 > 2078.76$ Then C2 (12/1)
If $CdO = 0$ & $Na_2O \leq 27414.98$ Then C2 (12/1)
If $Na_2O > 27414.98$ & $NiO \leq 58.42$ Then D1 (10/0)
If $NiO > 48.45$ & $CdO = 0$ & $BaO = 0$ & $Br_2O_7 \leq 53.6$ & $Fe_2O_3 \leq 12003.35$ & $ZnO \leq 149.31$ Then D1 (7/0)
Default class: LT D1 (2)

In parenthesis the number of covered cases is given followed by a number of errors each rule makes. These rules predict correctly 54 out of 68 test cases, using 10 features. Surprisingly, MnO selected by 1R has not been used at all. MnO is also selected at the top of SSV trees. Optimal SSV rules (pruning degree is optimized using internal crossvalidation) have slightly lower complexity, the tree has 11 nodes, including 6 leaves, and gives slightly better accuracy than C4.5 on the test set:

If $MnO < 1668.47$ & $ZrO_2 > 303.34$ Then C1
If $MnO < 1668.47$ & $ZrO_2 < 303.34$ & $TiO_2 < 76.235$, or
  $MnO > 1668.47$ & $Sb_2O_3 > 986.19$, or
  $MnO > 1668.47$ & $Sb_2O_3 < 986.19$ & $CaO < 79370$ Then C2
If $MnO > 1668.47$ & $Sb_2O_3 < 986.19$ & $CaO > 79370$, or
  $MnO < 1668.47$ & $ZrO_2 < 303.34$ & $TiO_2 > 76.235$ Then

D1

Since crossvalidation training is used in SSV in some runs a more complex solution, involving 12 rules that classify the training cases 100% correctly, is found. The increase of accuracy on the test set is rather small.

C. The Third Experiment (Broken Parts Data)

For the third experiment only samples with measurements on the broken parts were selected. There are 154 such instances in the database. The distribution of classes is:

1) C1, 32.47% (50 cases),
2) LT C2, 29.87% (46 cases),
3) LT D1, 37.66% (58 cases).

We divided them into 78-case training partition and took the remaining 76 as the test partition. The splitting was made in such a way that cases obtained from measurements on the same glass object are separated. Table (II) summarizes the best results obtained for this data.

TABLE III

RESULTS FOR THE BROKEN SIDE DATA.

System	Training %	Test %
1-NN Euclid (GM)	100	89.5
1-NN Manh (GM)	100	89.5
NaiveBayes (WEKA)	92.3	86.8
IncNet, 3 neurons (GM)	97.4	85.5
SVM, linear (GM)	94.9	81.6
MLP, 19 neurons (WEKA)	97.4	81.6
C4.5 - Rules (WEKA)	98.7	81.6
SSV, opt prune (GM)	84.6	77.6
C4.5 - Tree (WEKA)	93.6	77.6
1R (WEKA)	73.1	75.0

Good performance of the Naive Bayes should be noted The simplest rules obtained from 1R system for bucket = 7 (optimized in 5-fold crossvalidation on the training set) are:

If $MnO < 2134.61$ Then C1
If $MnO \in [2134.61, 9078.525)$ Then C2
If $MnO \geq 9078.525$ Then D1

These rules handle correctly 58 out of 78 training samples and 57 out of 76 test instances. Similar rules were found by the SSV tree with strong pruning.

The C4.5 decision tree produced 6 rules listed below:

If $ZrO_2 > 199.38$ & $CdO = 0$ Then C1 (19/0)
If $NiO \leq 62.23$ & $CaO \leq 114121.35$ Then C1 (6/0)
If $CuO \leq 5105.37$ & $MnO > 2546.77$ & $ZnO \leq 126.29$ Then C2 (15/0)
If $SnO_2 > 61.98$ & $Br_2O_7 \leq 64.08$ Then D1 (10/1)
If $Sb_2O_3 \leq 8246.11$ & $CuO \leq 2042.19$ & $Al_2O_3 > 11525.69$ Then D1 (20/0)
Default: C2 (8)

In the brackets the number of covered cases is given, followed by a number of errors each rule makes. These rules predict correctly 62 out of 76 test cases. SSV has also found 6 rules of similar complexity, although again using MnO as the ost important attribute.

IV. CONCLUSIONS AND FURTHER WORK

Numerical studies conducted here indicate that computational intelligence methods can be used for prediction of the membership of the glass samples of uncertain chronology to one of the main chronological periods. The most interesting result of this paper is however confirmation that place of the measurement (original surface or broken part of glass) has no influence on the results of analysis, and what follows prediction of membership of the sample to one of the chronological classes. This conclusion has been reached because the separate test on the surface and on the broken side of glass artifacts leads to similar classification accuracies by most classification systems used.

In the original database there is a significant proportion of the unlabeled samples. Assigning them to one of the chronological periods is of great importance for the archeologists. Predictions for the unlabeled cases made by machine learning systems should be confronted with archeological expertise. Methods that use the unsupervised learning procedures may help to improve the supervised classifiers, making such data useful to create better models.

With a growing database of glass samples it should be possible to create a rule based expert system, perhaps using fuzzy rules for estimation of probability, to help in determining chronology of new glass artifact, for example found in excavation sites. This paper is the first step towards such a system.

REFERENCES

[1] P. Wobrauschek, G. Halmetschlager, S. Zamini, C. Jakubonis, G. Trnka, M. Karwowski. *Energy-Dispersive X-Ray Fluorescence Analysis of Celtic Glasses.* In: Special Millennium Issue on Cultural Heritage (Ed. E.S. Lindgren), X-Ray Spectrometry 29, pp. 25-33, 2000.

[2] C. Jokubonis, P. Wobrauschek, S. Zamini, M. Karwowski, G. Trnka, P. Stadler. *Results of Quantitative Analysis of Celtic Glass Artifacts by Energy Dispersive X-ray Fluorescence Spectrometry.* Spectrochimica Acta Part B 58, pp.627-633, 2003.

[3] I.T. Nabney, C. Bishop, *NETLAB neural toolbox for Matlab*, Aston University, Birmingham, U.K. 2001.

[4] I.H. Witten, E. Frank, *Data Mining: Practical Machine Learning Tools and Techniques with Java Implementations.* Morgan Kaufmann Publishers, 2000.

[5] *SBL, Similarity-Based Learner*, software developed by K. Grudziński, Department of Informatics, Nicholaus Copernicus University, 1997-2002.

[6] *GhostMiner*, software developed by N. Jankowski, K. Grąbczewski, A. Naud and R. Adamczak, Department of Informatics, Nicholaus Copernicus University, 1997-2003. www.fqspl.com.pl/ghostminer/

[7] T. Hastie, R. Tibshirani, J. Friedman. *The Elements of Statistical Learning.* Springer Verlag 2001.

[8] K. Grąbczewski and W. Duch, *A General Purpose Separability Criterion for Classification Systems.* In: Proc. 4th Conf. on Neural Networks and Their Applications, Zakopane, Poland, pp. 203-208, May 1999.

[9] N. Jankowski and K. Grąbczewski, *Toward optimal SVM*, in preparation.

[10] N. Jankowski and V. Kadirkamanathan, *Statistical Control of RBF-like Networks for Classification.* In: 7th Int. Conf. on Artificial Neural Networks (ICANN 1997), Lausanne, Switzerland, Springer-Verlag, 385–390, 1997.