Machine learning from coronas using parametrization of images

Igor Kononenko, Matjaž Bevk, Aleksander Sadikov, Luka Šajn
University of Ljubljana
Faculty of Computer and Information Science
Ljubljana, Slovenia
e-mail: igor.kononenko@fri.uni-lj.si

Abstract

We were interested to develop an algorithm for detection of coronas of people in altered states of consciousness (two-classes problem). Such coronas are known to have rings (double coronas), special branch-like structure of streamers and/or curious spots. We used several approaches to parametrization of images and various machine learning algorithms. We compared results of computer algorithms with the human expert's accuracy. Results show that computer algorithms can achieve the same or even better accuracy than that of human experts.

1. Introduction

Recently developed technology by Korotkov (1998) from Technical University in St.Petersburg, based on the Kirlian effect, for recording the human bio-electromagnetic field (aura) using the Gas Discharge Visualization (GDV) technique provides potentially useful information about the biophysical and/or psychical state of the recorded person. In order to make the unbiased decisions about the state of the person we want to be able to develop the computer algorithm for extracting information/describing/classifying/making decisions about the state of the person from the recorded coronas of fingertips.

The aim of our study is to differentiate 6 types of coronas, 3 types in normal state of consciousness: **Ia, Ib, Ic** (pictures were recorded with single GDV camera in Ljubljana, all with the same settings of parameters, classification into 3 types was done manually):

- **Ia** – harmonious energy state (120 coronas)
- **Ib** – non-homogenous but still energetically full (93 coronas)
- **Ic** – energetically poor (76 coronas)

and 3 types in altered states of consciousness (pictures obtained from dr. Korotkov, recorded by different GDV cameras with different settings of parameters and pictures were not normalized – they were of variable size):

- **Rings** – double coronas (we added 7 pictures of double coronas recorded in Ljubljana) (90 coronas)
- **Branches** – long streamers branching in various directions (74 coronas)
- **Spots** – unusual spots (51 coronas)

Our aim is to differentiate normal from altered state of consciousness (2 classes) and to differentiate among all 6 types of coronas (6 classes). Figures 1a and b provide example coronas for each type.

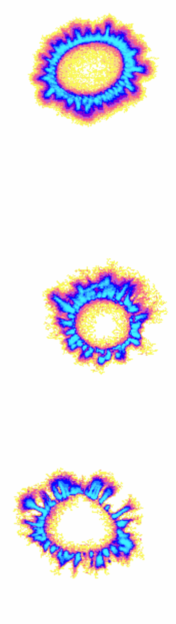

Figure 1a: Example coronas for Types Ia, Ib and Ic– normal state of consciousness

2. The methodology

We first had to preprocess all the pictures so that all were of equal size (320 x 240). We then described the pictures with various sets of numerical parameters (attributes) with five different parametrization algorithms:

a) IP (Image Processor – 22 attributes) (Bevk and Kononenko, 2002),
b) PCA (Principal Component Analysis) (Turk and Pentland, 1991),
c) Association Rules (Bevk, 2003),
d) GDV Assistant with some basic GDV parameters (Korotkov, 1998; Sadikov, 2002),
e) GDV Assistant with additional parameters (Sadikov, 2002).

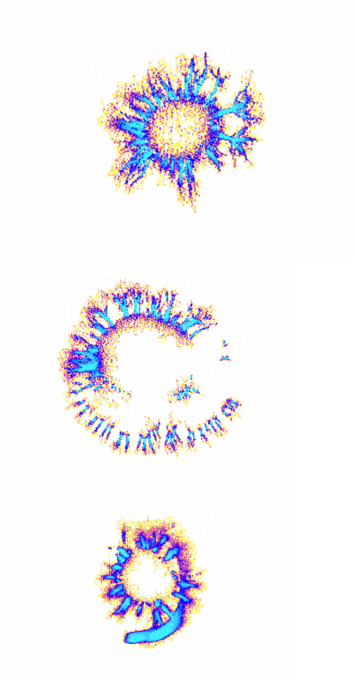

Figure 1b: Example coronas for Types Branches, Rings and Spots– altered states of consciousness

Therefore we had available 5 different learning sets for two-classes problem: altered (one of Rings, Spots, and Branches) versus non- altered (one of Ia, Ib, Ic) state of consciousness. Some of the sets were used also as six-classes problems (differentiating among all six different types of coronas).

We tried to solve some of the above classification tasks by using various machine learning algorithms as implemented in Weka system (Witten and Frank, 2000):

- Quinlan's (1993) C4.5 algorithm for generating decision trees;
- K-nearest neighbor classifier by Aha, D., and D. Kibler (1991);
- Simple Kernel Density classifier;
- Naïve Bayesian classifier using estimator classes: Numeric estimator precision values are chosen based on analysis of the training data. For this reason, the classifier is not an Updateable Classifier (which in typical usage are initialized with zero training instances, see (John and Langley, 1995));
- SMO implements John C. Platt's sequential minimal optimization algorithm for training a support vector classifier using polynomial kernels. It transforms the output of SVM into probabilities by applying a standard sigmoid function that is not fitted to the data. This implementation globally replaces all missing values and transforms nominal attributes into binary ones (see Platt, 1998; Keerthi et al., 2001);
- Neural networks: standard multilayared feedforward neural network with backpropagation of errors learning mechanism (Rumelhart et al., 1986).

SMO algorithm can be used only for two-classes problems, while the other algorithms can be used on two-classes and on six-classes problems.

3. Results

3.1 Results for C4.5

In the first experiment we tried to learn decision trees from five different descriptions of data, as returned by different parametrization algorithms, for two-classes problem. We used Quinlan's C4.5 algorithm. For testing we used the standard 10-fold cross validation. Results in Table 1 show that GDV Assistant achieves best results, which was also expected, although we expected larger advantage over other parametrization algorithms. Additional attributes for description of different statistics of fragments did not provide any improvement, which is somehow disappointing.

	Number of attributes	Classification error	Standard error	Default error
IP	22	19.8 %	0.9 %	43,0 %
PCA	15	29.2 %	1.8 %	43,0 %
Assoc.rules	44	20.0 %	1.6 %	43,0 %
GDV Assist.	17	18.6 %	1.5 %	43,0 %
GDV Assist with add. atts	27	18.3 %	1.5 %	43,0 %

Table 1: Classification error of C4.5 on five different descriptions of coronas for two-class problem

3.2 Results of a human expert

In order to get a better feeling about how good are the above results in comparison to humans, we tested a human expert on one fold (51 testing instances) and compared his result with the accuracy of C4.5 (using the parametrization with

Associative rules) on the same testing set. On the two-classes problem the human expert and C4.5 achieved the same classification error of 23.5%. It seems that C4.5 is biased towards classifying more coronas as normal and therefore coronas for the altered state of consciousness are poorly classified. On the other hand the human expert does not have such bias and the misclassifications are more evenly distributed between two classes.

On the six-classes problem the human expert had classification error of 39,2%. C4.5 was significantly worse with 54.9%. Again it seems that C4.5 is biased towards classifying more coronas as normal and therefore coronas for the altered state of consciousness are poorly classified. On the other hand the human expert does not have such bias and the misclassifications are more evenly distributed among normal an altered states of consciousness.

The easiest types of coronas for classification are Ia (normal) and Rings (altered). The most difficult type of coronas seems to be Branches (altered), which is by human expert most often confused with Ib (normal) and Spots (altered), and by C4.5 with Ic (altered).

3.3 Results of other machine learning algorithms

	C4.5	Naïve Bayes	K-NN	Kernel Density	SMO	Neural networks
IP	80.2 %	82.7 %	78.7%	76.2 %	84.7 %	83.1 %
PCA*	72.8 %	74.2 %	71.2 %	66.0 %	74.3 %	74.3 %
Assoc.rules*	80.9 %	75.1 %	80.0 %	76.8 %	83.8 %	83.9 %
GDV Assist	81.4 %	81.2 %	74.5 %	65.7 %	80.6 %	76.9 %
GDV Assist with add. atts	81.7 %	80.9 %	72.6 %	64.6 %	80.5 %	71.9 %

Table 2: Classification accuracy of five machine learning algorithms on three different descriptions of coronas for two-classes problem (* we give best results over ten different pre-training subsets of images for Association rules and PCA parametrization algorithms)

We tried also the other machine learning algorithms on all the data sets. For testing we used the standard 10-fold cross validation. The results are given in Tables 2 and 3. Parametrization algorithms PCA and Association rules need a small subset of images for defining the attributes (this subset, called pre-training subset, is subtracted from the training set of images). We run each of these two algorithms on ten different pre-training subsets of images, so that we obtained for each of them ten different parametrizations. All machine learning algorithms were run on all of them and in tables we give the best results (the highest accuracy among ten average accuracies obtained from cross validations over all ten parametrizations). Note also that SMO can be used only for two-classes problems, therefore it is omitted from Table 3.

On the two-classes problem SMO (based on SVM machine learning algorithm) achieved the best results: accuracy was up to 85%. Neural networks were rather close with accuracy up to 84% using Associative rules for parametrization. The worse was Kernel Density algorithm, while the others achieved comparable results.

GDV Assistant, Image Processor with statistical parametrization, and parametrization based on Association rules provide comparably good description of images, while the Principal Component Analysis provides the worst parametrization. This is somehow disappointing as we expected PCA to be more appropriate for describing coronas than algorithms, which are designed specially for textures. However, it is well known that PCA requires normalized images (coronas should have been all of approximately equal size and centered, pictures should all be of equal level of brightness) which was not the case in our study.

	C4.5	Naïve Bayes	K-NN	Kernel Density	Neural networks
IP	60.2 %	55.4 %	56.8 %	55.5 %	65.2 %
PCA*	47.7 %	50.7 %	50.2 %	53.8 %	52.1 %
Assoc.rules*	50.9 %	37.5 %	51.5 %	52.5 %	61.6 %
GDV Assist	55.0 %	54.2 %	51.6 %	49.4 %	59.0 %
GDV Assist with add atts	54.1 %	51.2 %	48.8 %	48.5 %	54.6 %

Table 3: Classification accuracy of five machine learning algorithms on three different descriptions of coronas for six-classes problem (* we give best results over ten different pre-training subsets of images for Association rules and PCA parametrization algorithms)

On the six-classes problem neural networks seems to perform best and achieved accuracy up to 65% using statistical parametrization of images. The other algorithms (without SMO, which cannot deal with more than two classes) achieved lower accuracy. Image Processor with statistical parametrization shows clear advantage over the other parametrization methods, which give comparable quality of image descriptions.

4. Conclusions and future work

The most competitive machine learning algorithms for our problem seem to be neural networks and Support vector machines (SVM incorporated in SMO algorithm). The worst seems to be the Kernel Density classifier.

Among parametrization techniques GDV Assistant, Image Processor (statistical approach) and Associative rules (symbolic approach) achieved similar quality of image descriptions, while the Principal Component Analysis was the worse. GDV Assistant seems to be a promising approach, however further tests and further improvements are necessary.

The best results (classification accuracy up to 85%) in the two-classes problem were achieved by SMO algorithm, based on Support vector machines, and using statistical parametrization of images. This result is better than that of a human expert, who achieved 77% of classification accuracy. In the six-classes problem, the best results (classification accuracy up to 65%) were achieved by neural networks with statistical parametrization of images. This result also outperforms that of the human expert. Those results indicate that computer algorithms can achieve the same or even better accuracy than human experts.

In future we plan to:
- add to GDV Assistant some more possibly informative parameters for this problem,
- we plan to try some other machine learning algorithms (other variants of decision trees and Naïve Bayes),
- test more human experts to see how accurate the manual classification can be,
- combine various parametrization techniques in order to extract possibly different and useful information from different sets of parameters and preprocess is using various feature subset selection approaches,
- combine the decisions of different classifiers in order to improve their reliability and try also other approaches to improving the classification accuracy, such as bagging, boosting, stacking, and transduction,
- we plan to collect additional images of coronas in order to increase the number of training/testing instances and therefore to improve the reliability of classifiers.

Acknowledgments

We thank dr. Konstantin Korotkov for providing the coronas of fingertips of people in altered states of consciousness. We thank also Vida Korenc for preprocessing the images and Ivan Vidmar for implementing additional parameters in GDV Assistant and for performing preliminary experiments with C4.5.

References

Aha, D., and D. Kibler (1991) "Instance-based learning algorithms", Machine Learning, vol.6, pp. 37-66.

Bevk M. (2003) Texture Analysis with Machine Learning, M.Sc. Thesis, University of Ljubljana, Faculty of Computer and Information Science, Ljubljana, Slovenia. (in Slovene)

M. Bevk and I. Kononenko (2002) A statistical approach to texture description: A preliminary study. In ICML-2002 Workshop on Machine Learning in Computer Vision, pages 39-48, Sydney, Australia, 2002.

G. H. John and P. Langley (1995). Estimating Continuous Distributions in Bayesian Classifiers. Proceedings of the Eleventh Conference on Uncertainty in Artificial Intelligence. pp. 338-345. Morgan Kaufmann, San Mateo.

S.S. Keerthi, S.K. Shevade, C. Bhattacharyya, K.R.K. Murthy (2001). Improvements to Platt's SMO Algorithm for SVM Classifier Design. Neural Computation, 13(3), pp 637-649, 2001.

Korotkov, K. (1998) *Aura and Consciousness*, St.Petersburg, Russia: State Editing & Publishing Unit "Kultura".

J. Platt (1998). Fast Training of Support Vector Machines using Sequential Minimal Optimization. Advances in Kernel Methods - Support Vector Learning, B. Schölkopf, C. Burges, and A. Smola, eds., MIT Press.

W.H. Press, B.P. Flannery, S.A. Teukolsky, and W.T. Vetterling (1992) *Numerical Recipes: The Art of Scientific Computing*. Cambridge University Press, Cambridge (UK) and New York, 2nd edition.

J.R. Quinlan (1993) C4.5 Programs for Machine Learning, Morgan Kaufmann.

D.E. Rumelhart, G.E. Hinton, R.J. Williams (1986) Learning internal representations by error propagation. In: Rumelhart D.E. and McClelland J.L. (eds.) Parallel Distributed Processing, Vol. 1: Foundations. Cambridge: MIT Press.

A. Sadikov (2002) Computer visualization, parameterization and analysis of images of electrical gas discharge (in Slovene), M.Sc. Thesis, University of Ljubljana, 2002.

L. Sirovich and M. Kirby (1987) A low-dimensional procedure for the characterisation of human faces. Journal of the Optical Society of America, pages 519-524.

M. Turk and A. Pentland (1991) Eigenfaces for recognition. *Journal of Cognitive Neuroscience*, pages 71-86.

I. H. Witten, E. Frank (2000) Data mining: Practical machine learning tools and techniques with Java implementations, Morgan Kaufman.