Classification of Boar Sperm Head Images using Learning Vector Quantization

Biehl, Michael; Pasma, Piter; Pijl, Marten; Sánchez, Lidia; Petkov, Nicolai

Published in:
Proc. European Symposium on Artificial Neural Networks

Publication date:
2006

Classification of Boar Sperm Head Images using Learning Vector Quantization

Michael Biehl[1], Piter Pasma[1], Marten Pijl[1], Lidia Sánchez[2] and Nicolai Petkov[1]

1- Rijksuniversiteit Groningen - Mathematics and Computing Science
P.O. Box 800, NL-9700 AV Groningen - The Netherlands

2- University of León - Department of Electrical and Electronics Engineering
Campus de Vegazana s/n, 24071 León - Spain

Abstract. We apply Learning Vector Quantization (LVQ) in automated boar semen quality assessment. The classification of single boar sperm heads into healthy (*normal*) and *non-normal* ones is based on grey-scale microscopic images only. Sample data was classified by veterinary experts and is used for training a system with a number of prototypes for each class. We apply as training schemes Kohonen's LVQ1 and the variants Generalized LVQ (GLVQ) and Generalized Relevance LVQ (GRLVQ). We compare their performance and study the influence of the employed metric.

1 Introduction

Semen quality assessment constitutes an important problem in fertility studies. Existing, relatively costly methods rely on staining or motility measurements. Mostly they have originally been developed for the analysis of human semen and clinical purposes. Here, the aim is a fast and relatively cheap reliable method for the inspection of boar semen samples on the basis of microscopic images.

We apply methods of prototype based clustering and classification. Learning vector quantization (LVQ) as originally proposed by Kohonen is widely used in a variety of areas due to its flexibility and conceptual simplicity [1, 4]. LVQ provides a particularly intuitive and clear approach: the classification of data is based on a set of so-called prototype vectors and a feature vector is assigned to the class represented by the *closest* prototype.

Frequently, distances are evaluated in terms of simple Euclidean metrics, but alternative measures can easily be incorporated into the standard LVQ schemes. The choice of an appropriate distance measure can indeed be crucial in practical problems. Recent variations of LVQ have been suggested in which an adaptive metric is modified in the course of learning.

One of the most attractive features of LVQ learning is that the parameterization in terms of prototype vectors allows for an immediate interpretation of the classifier. Prototypes are defined in the same space as the data; they are, for instance, images themselves and provide direct information about the achieved clustering and the features that it is based on. This is in contrast to other methods, including feed-forward neural networks and the support vector machine which in general cannot be interpreted as easily. The paper is organized as follows: in the next section we outline the problem and summarize how the

data has been acquired and pre-processed. In section 3, we describe the LVQ classifier and training algorithms. The results are presented and discussed in section 4 and we conclude with a brief summary in 5.

2 Classification problem and data acquisition

Computer assisted analysis of semen fertility has been used to save costs and improve quality in artificial insemination. Sperm analysis is a complex task which allows to identify the males with the best reproductive features and control the fertility of the samples. Some approaches have used image analysis techniques to detect shape abnormalities in sperm heads [6] or classify subpopulations in a sample. However, only very few methods use the information of the acrosome integrity or the intracellular distribution to evaluate the sperm quality [7, 8].

Several commercial systems have been developed, which were designed specifically for human sperm cells and have been adapted to other species later. They do not consider the intracellular density distribution which provides information about the cell status.

Veterinary experts hypothesize that there are certain patterns of cytoplasm density distributions which can be correlated with alive and dead sperm cells. Techniques such as fluorescence staining protocols mark sperm heads depending on if they are alive of dead. However, these methods are expensive and require more work in the sample preparation. The goal of this work is to classify the sperm head images without using staining techniques in order to speed up the process and reduce the costs.

The 1360 images used in this project are obtained from boar semen samples by use of a phase-contrast microscope. The first processing and segmentation steps provide aligned images of isolated sperm heads as grey-level distributions on a black (*empty*) background. For details of acquisition and pre-processing see [7].

Sperm heads typically display an oval shape which we approximate by an ellipse. We determine its principal axes and align all images. Using nearest-neighbor interpolation, we re-scale all head images to the same aspect ratio of 19 x 35 pixels. Finally, we consider an array of grey-levels which lie inside the ellipse with main axes 19 and 35 pixels.

Intracellular intensities in heads which were labelled as *normal* by veterinary experts display three areas: a darker region corresponding to the post nucleus cap where the tail develops from, an intermediate light area and the acrosome which is covering the nucleus region, see Fig. 1 for example images. As the illumination of the samples under the microscope is controlled manually, brightness and contrast can vary significantly across different images. To cope with this problem, we perform a linear transformation which results in the same mean, $\mu = 100$, and standard deviation, $\sigma = 8$, of the grey-levels in all images.

We represent example images as vectors $\boldsymbol{\xi}^\mu \in I\!R^N$ where the $N = 35 \times 19 = 665$ components correspond to a rectangle. All pixels outside the above mentioned ellipsoidal regions are the same in every $\boldsymbol{\xi}^\mu$ and will play no role in the LVQ training and classification performance. Due to variations in shape, also the ellipses can contain pixels or entire regions which do not correspond

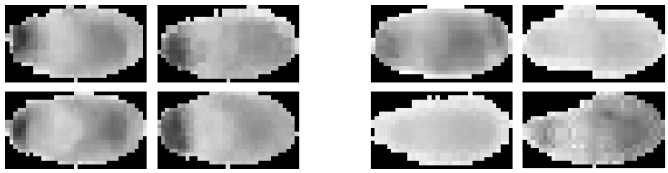


Fig. 1: Examples of sperm heads which were labelled as *normal* (left panel) or *non-normal* (right panel), respectively. For clarity, background pixels are displayed black while for the classification they are set to the overall mean value of all non-background pixels, see the discussion in section 2.

to intracellular densities but are part of the *black background* in Fig. 1. In [7], such pixels were explicitly excluded when comparing different heads, which essentially disregards differences in shape. Here, we have chosen to replace them by a constant value, i.e. the total mean of non-background grey-scales over all example images. Thus, we are taking into account deformations but avoid artificially large differences between corresponding pixels. By choice of the constant one could vary the emphasis that is put on shape, the effect of this will be studied in a forthcoming project.

3 LVQ system and training algorithms

Training is based on the repeated random sequential presentation of example images drawn from a given (sub-) set of data $I\!D = \{\boldsymbol{\xi}^\mu, S_T^\mu\}_{\mu=1}^P$. Here the class membership label as provided by the veterinary experts is

$$S_T^\mu = +1 \;\text{ if } \boldsymbol{\xi}^\mu \text{ is marked as } \textit{normal} \quad \text{and } S_T^\mu = -1 \;\text{ else.} \tag{1}$$

In the set of M prototypes $\{\mathbf{w}^1, \mathbf{w}^2, \ldots, \mathbf{w}^M\}$ a vector $\mathbf{w}^k$ is supposed to represent data from class $S^k \in \{-1, +1\}$. Upon presentation of the pair $(\boldsymbol{\xi}^\mu, S_T^\mu)$ at time step t, we evaluate the distance of $\boldsymbol{\xi}^\mu$ from the current prototypes $\mathbf{w}^j(t)$. In the following we will use three different distance measures $d(j, \mu)$:

a) (squared) Euclidean: $\quad d_2(j,\mu) = \left(\boldsymbol{\xi}^\mu - \mathbf{w}^j(t)\right)^2 = \sum_{i=1}^N \left(\xi_i^\mu - w_i^j(t)\right)^2$

b) generalized (squared) Euclidean: $\quad d_{2,\lambda}(j,\mu) = \sum_{i=1}^N \lambda_i \left(\xi_i^\mu - w_i^j(t)\right)^2$
with the positive and normalized *relevances* $\lambda_i > 0$, $\sum_i \lambda_i^2 = 1$

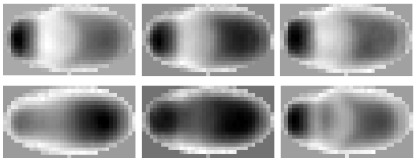
c) L_1 metric: $\quad d_1(j,\mu) = \sum_{i=1}^N \left|\xi_i^\mu - w_i^j(t)\right|$.

In any case we identify the minimal distances $d(j, \mu)$ among prototypes with class labels $S_T^j = +1$ and $S_T^j = -1$ separately and use the notation

$$\begin{cases} \mathbf{w}^J(t) & \textit{(correct winner)} \quad \text{with } d(J,\ \mu) = \min\limits_k \left\{d(k,\mu) | S^k = +S_T^\mu\right\} \\ \mathbf{w}^K(t) & \textit{(wrong winner)} \quad \text{with } d(K,\mu) = \min\limits_k \left\{d(k,\mu) | S^k = -S_T^\mu\right\}. \end{cases} \tag{2}$$

Further, the *total winner* is denoted as $\mathbf{w}^L(t) = \begin{cases} \mathbf{w}^J(t) & \text{if } d(J,\mu) < d(K,\mu) \\ \mathbf{w}^K(t) & \text{else.} \end{cases}$

We consider the following training schemes:



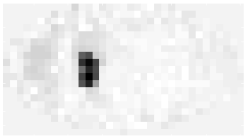


Fig. 2: Left panel: The upper (lower) row displays prototypes in LVQ1 for *normal (non-normal)* heads, respectively. Right panel: Relevances as obtained by GRLVQ: the classification is based on very few components with considerable λ_i, corresponding to the dark region in the figure.

a) **LVQ1**
At each learning step, only the total winner $\mathbf{w}^L(t)$ is updated:

$$\mathbf{w}^L(t+1) = \mathbf{w}^L(t) + \eta(t)[S_T^\mu S^L]\left(\boldsymbol{\xi}^\mu - \mathbf{w}^L(t)\right) \tag{3}$$

The update is towards (away from) the input $\boldsymbol{\xi}^\mu$ if the class label of the winning prototype and that of the example agree (disagree). In LVQ1 we have considered the use of the distance measures $d_2(j,\mu)$ and $d_1(j,\mu)$.

b) **GLVQ**
In this learning scheme we restrict our implementation to the Euclidean distance measure $d(j,\mu) = d_2(j,\mu)$. Both winning prototypes, $\mathbf{w}^J$ and $\mathbf{w}^K$, are modified. Defining $z^\mu = (d(J,\mu) - d(K,\mu))/\left(d(J,\mu) + d(K,\mu)\right)$ the update of $\mathbf{w}^J$ can be written as

$$\mathbf{w}^J(t+1) = \mathbf{w}^J(t) + \eta(t)[S_T^\mu S^J] f'\left(z^\mu\right) \frac{d(K,\mu)}{(d(J,\mu)+d(K,\mu))^2}\left(\boldsymbol{\xi}^\mu - \mathbf{w}^J(t)\right) \tag{4}$$

and that of $\mathbf{w}^K$ is of the same form with indices K and J exchanged. This corresponds to a step in the direction of the negative gradient of the function $f(z^\mu)$ w.r.t. $\mathbf{w}^K$ and $\mathbf{w}^L$. The function f has been set to $f(x) = (1+e^{-x})^{-1}$ in [2, 3]. For the limited range of possible arguments $-1 \le z^\mu \le 1$ the sigmoidal is very close to linear, however, and we replace it by the simpler $f(x) = x, f'(x) = 1$ in the following.

c) **GRLVQ**
In GRLVQ as suggested in [3], the distance measure $d(j,\mu) = d_{2,\lambda}(j,\mu)$ is used and the cost function $f(z^\mu) = z^\mu$ of GLVQ is re-defined accordingly. Both prototypes $\mathbf{w}^J(t)$, $\mathbf{w}^K(t)$ and also the relevances λ_i are updated in the direction of the respective negative gradient, see [3] for details.

In all considered algorithms the learning rate should gradually decrease in the course of learning. We have implemented a schedule of the form

$$\eta(t) = \eta_o \text{ for } \tau < \tau_o \quad \text{and} \quad \eta(t) = \eta_o/[1 + b(\tau - \tau_o)] \text{ for } \tau \ge \tau_o.$$

Here, τ counts the number of randomly shuffled sweeps through the actual training set. In each training process we have performed 80 such sweeps and used the values $\tau_o = 10$, $\eta_o = 0.1$, and $b = 0.3$.

4 Results

In order to evaluate the quality of the obtained classification we have performed ten-fold cross-validation: We split the set of 1360 available data into ten disjoint subsets $\mathbb{D}_i$ of equal size. For a given number of prototypes each of ten identically designed LVQ systems ($n = 1, 2, \ldots 10$) was trained from $\cup_{i \neq n} \mathbb{D}_i$, while $\mathbb{D}_n$ served as a test set. The following table summarizes some of our results:

# of prototypes:	3 / 3 (*normal*/*non-normal*)	1 / 7
LVQ1 (L_1 measure)	$\epsilon_{test} = 0.232$ ($\sigma = 0.037$)	$\epsilon_{test} = 0.215$ ($\sigma = 0.058$)
LVQ1 (Euclidean)	$\epsilon_{test} = 0.186$ ($\sigma = 0.040$)	$\epsilon_{test} = 0.184$ ($\sigma = 0.045$)
GLVQ (Euclidean)	$\epsilon_{test} = 0.244$ ($\sigma = 0.041$)	$\epsilon_{test} = 0.236$ ($\sigma = 0.038$)
GRLVQ (Euclidean)	$\epsilon_{test} = 0.185$ ($\sigma = 0.035$)	$\epsilon_{test} = 0.183$ ($\sigma = 0.037$)

Here ϵ_{test} is the mean test error and σ denotes its standard deviation over the ten (statistically dependent) training sets. Note that a training error of 0.20 comparable with the above test error was achieved for the same data set using a single normal cell prototype and a distance threshold [7, 8]. We are not aware of other published results for boar sperm head classification based on microscopic images only.

As a first result we note that the performance depends only weakly on the number of prototypes. Relatively low test errors can be achieved by using a single prototype for the class of *normal* heads while the *non-normal* sperms are represented by several prototypes. We observe only weak variations and no indication of significant over–fitting in the considered range of $1-10$ *normal* and $3-16$ *non-normal* prototypes. We have studied the performance of LVQ1 using the squared Euclidean distance as well as the L_1-metric. In all considered cases we have found a lower classification error when using the Euclidean measure and, hence, we restrict our comparison of different algorithms to this case.

An important observation is that the performance of GLVQ is inferior to that of LVQ1 (with Euclidean distances). One attractive feature of GLVQ is that it can be interpreted as a gradient descent with respect to a plausible cost function. However, the relation of the latter with the goal of good generalization is not obvious. Indeed, we observe that, here, training and test error typically increase in the course of GLVQ training while the associated cost function decreases.

By introducing adaptive relevances, the good performance of LVQ1 is recovered in GRLVQ. It is interesting to note that only very few relevances λ_i remain significantly different from zero in the training process. Fig. 2 (right panel) displays a typical vector of relevances in grey-scale. GRLVQ selects mainly those pixels which correspond to the pronounced light region in *normal* head images. Apparently, the achieved performance could be obtained from a comparison of these regions only, while the acrosome and post nucleus cap densities are essentially disregarded. This seems to indicate that the resolution of the obtained images might be insufficient to exploit the latter information.

5 Summary and Outlook

We have applied several LVQ schemes in the classification of boar sperm head

images. Unlike other more sophisticated and more expensive approaches, we aim at a classification which is based on grey-scale microscopic images only. We are aware of only two earlier publications [7, 8] which report results based on similar data and we find that plausible LVQ algorithms yield comparable, if not superior performance. It is very well feasible that a more careful fine-tuning of the LVQ algorithms, e.g. with respect to the learning rate schedule, will yield an even lower classification error.

Cross validation estimates of ϵ_{test} display a rather weak dependence on the number of employed prototypes. When comparing the use of L_1- *vs.* Euclidean distances in LVQ1, we find the latter to be superior and restrict the further analysis to this case. While the heuristic LVQ1 yields relatively low test error, here, the cost function based GLVQ is clearly inferior. Better performance is recovered when introducing adaptive relevance factors in the frame of GRLVQ.

The training of relevance factors reveals insight into what features the classification is based on. We find that a test error of about 0.20 can be achieved by comparing a small characteristic region of the grey-scale data. The latter corresponds to a pronounced low density region in images of *normal* sperm heads. While acrosome density is believed to carry relevant information, they are essentially disregarded here. In forthcoming investigations we will aim at further optimization of the training process and LVQ architectures. For now, it remains an open question if the data in its current form is sufficient to achieve significantly better classification schemes.

Finally, we would like to point out that the goal in semen quality assessment is not necessarily a low error rate with respect to single image classification. The aim is to estimate with high precision the fraction of *non-normal* heads in a sample of unknown composition [8]. The design of a classifier appropriate for this purpose will be aimed at in a forthcoming project.

References

[1] *Bibliography on the Self-Organizing Map (SOM) and Learning Vector Quantization (LVQ)*, Neural Networks Research Centre, Helsinki University of Technology, 2002.

[2] A.S. Sato and K. Yamada, *Generalized learning vector quantization*. In: G. Tesauro, D. Touretzky, and T. Leen, eds., *Advances in Neural Information Processing Systems 7* (MIT Press, 1995)

[3] B. Hammer and T. Villmann, *Generalized relevance learning vector quantization*, Neural Networks 15: 1059-1068, 2002.

[4] T. Kohonen, *Self-organizing maps*, Springer, Berlin, 1995.

[5] L. Thurston, W. Holt, and P. Watson. Post-thaw functional status of boar spermatozoa cryopreserved using three rate freezers: a comparison. *Theriogenology*, 60:101–113, 2003.

[6] M. Hirai et, A. Boersma, A. Hoeflich, E. Wolf, J. Foll, T.R. Aumuller, and J. Braun. Objectively measured sperm motility and sperm head morphometry in boars (Sus scrofa): relation to fertility and seminal plasma growth factors. *J. Androl*, 22:104–110, 2001.

[7] L. Sánchez, N. Petkov, and E. Alegre. Statistical Approach to Boar Semen Head Classification Based on Intracellular Intensity Distribution. In A. Gagalowicz and W. Philips, editors, *Proceedings of 11th International Conference, CAIP 2005*, Lecture Notes in Computer Science, Vol. 3691, pages 88–95. Springer-Verlag, 2005.

[8] L. Sánchez, N. Petkov, and E. Alegre. Classification of boar spermatozoid head images using a model intracellular density distribution. In M. Lazo and A. Sanfeliu, editors, *Progress in Pattern Recognition, Image Analysis and Applications: Proc. 10th Iberoamerican Congress on Pattern Recognition, CIARP 2005*, pages 154-160. Springer-Verlag, 2005.