Sparse Bayesian Feature Selection Applied to Intestinal Motility Analysis

L. Igual[1], S. Seguí[1], J. Vitrià[1], F. Azpiroz[2] and P. Radeva[1]
[1]Computer Vision Center and Universidad Autónoma de Barcelona, Bellaterra, Spain
[2]Hospital de "Vall d'Hebron", Barcelona, Spain

Abstract

In this paper we propose an innovative automatic system for diagnosing severe intestinal motility dysfunctions based on a joint feature selection and classifier learning. A novel acquisition technology, the Wireless Capsule Video Endoscopy, is used as source of data and visual information is exploited to characterize intestinal motility dysfunctions. In particular, the method automatically selects the most relevant visual features among a set of them defined by medical experts and, simultaneously, carries out the classification training. Thus, the dimension of the input data is reduced without any accuracy loss. To assess the accuracy of the proposed method we use the results obtained with the currently applied invasive diagnosis test, the manometry.

1 Introduction

In many medical classification problems the definition of the data attributes is a hard process which leads to a large set of features with different relevance. In these cases, feature selection approaches can be used to preserve the most informative feature subset for the classification task. Feature selection methods are useful when the input data dimension is large compared with the number of samples of the data set and it has been shown that they improve generalization when irrelevant features are present [1]. According to [1], feature selection approaches are divided into filters, wrappers and embedded approaches. The most common ones are filters which act as a preprocessing step independent of the final classifier. In contrast, wrappers consider the classifier as a black box. Finally, embedded approaches simultaneously determine features and classifier parameters during the training process. There are specifically tailored methods for feature selection in non-parametric classifier learning. Guyon et al. [1] proposed a feature ranking method for Support Vector Machines (SVM) where the selection is performed by Recursive Feature Elimination while Weston et al. [2] used the gradient descent method over an upper bound of the leave-on-out error for selecting the optimal subspace. Nevertheless, there exist only few embedded methods addressing feature selection in connection with non-parametric classifiers up to now. Recently, Krishnapuram et al. developed a *joint classifier and feature optimization* method for kernel classifiers [3]. The method shows promising results in high dimensional data sets. Nevertheless, the resulting kernel classifier model is learned using the Expectation Maximization algorithm and the computational complexity becomes unpractical for large training data sets.

In [4] a method for joint feature selection and classifier learning (JFSCL) using a sparse Bayesian approach is presented. This method optimizes a global loss function that includes a term associated with the empirical loss and another one representing a feature selection and regularization constraint on the classifier parameters. To minimize this function a recently proposed technique, the Boosted Lasso algorithm is used. It follows the regularization path of the empirical risk associated with the loss function. The algorithm is developed for a non-parametrical classification method, the Relevance Vector Machine (RVM).

In this paper, we propose to apply this JFSCL method to a medical imaging classification problem: the diagnosis of severe intestinal motility dysfunctions. Currently, the most extended diagnosis test for small bowel motility disorders is intestinal manometry [5], which measure the contractile activity as variations in pressure. This technique is highly invasive and requires hospitalization of the patient and monitorization of the whole process by medical staff. A recently appeared alternative acquisition method is the Wireless Capsule Video Endoscopy which consists of a capsule with a camera, a battery and a set of led lamps for illumination. This capsule is swallowed by the patient, emitting a radio frequency signal which is received and stored in an external device. The result is a video movie which records the "travel" of the capsule along the intestinal tract. This novel technique is much less invasive than the manometry. Moreover, novel approaches using this new video endoscopy data have been developed and successfully applied to detect some intestinal affections such as cancer, bowel Crohn's disease and small bowel ulcers [6]. However, as far as we know, WCVE has not been used to deal with motility diseases, being this a challenging field of research. The aim of this work is to automatically detect intestinal motility dysfunctions using WCVE as a source of data.

We formulate this problem as a binary classification task which requires a previous process to define the feature space. Note that this feature space is as large as we want, since the provided visual information is very rich. Thus, we are changing the one-dimensional feature space of the manometry to a new space with many visual features. Experts carry out the analysis of the intestinal elements and events present in the zone of interest of the downloaded video, and decide which ones are physiologically the most interesting ones. We assign a label to each relevant basic intestinal characteristic. After some videos are labelled, the experts study the resulting patterns of the intestinal activity based on the labels, leading to the definition of the video features which better characterize the intestinal behavior of subjects with some dysfunctions (we call them patients) and distinguish it from the normal activity of healthy subjects. Once the features are defined, our goal is to select the most relevant ones in terms of motility diseases diagnosis. The JFSCL method is used to perform the feature selection and the classification of a new test set.

The paper is organized as follows: in Section 2 we describe in detail the characteristics of the new video data obtained by capsule endoscopy and we introduce the employed methodology. In Section 3, we present the experimental results, and finally, we expose the conclusions.

2 Material and Method

The video frames recorded by the wireless endoscopic capsule have a resolution of 256×256 pixels, with a circular field of view of 240 pixels of diameter and spanning 140 degrees. In Figure 1 we display a frame, where the gut wall and lumen is visualized.

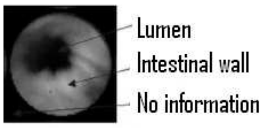

Figure 1: Example of intestinal video image

Medical experts concentrate in a portion of the video to study the motility dysfunctions. They are in charge of the assignment of frame labels and the definition of video features. The defined labels are the following: *turbid*, *tunnel*, *static*, *level of wrinkle presence*, *phasic contractions* and *non-occlusive contractions*. The turbid is food in digestion or intestinal juices. A tunnel is a sequence of frames where the lumen is static for a long period of time. Static frames appears when the camera has a null apparent motion and the visualized frames are almost the same. The wrinkle star pattern is an omnipresent characteristic of the sustained contractions. In the frames where the wrinkle pattern appears we observe strong edges of the folded intestinal wall, distributed in a radial way around the intestinal lumen. Phasic contractions are the result of muscular stimulation produced by the nervous system. In occlusive contractions the lumen in the central frame is completely closed, whereas in non-occlusive contractions the lumen never appears completely closed. Sustained contractions can be visualized as a continuous closing of the intestinal lumen. Figure 2 display some examples of these labels.

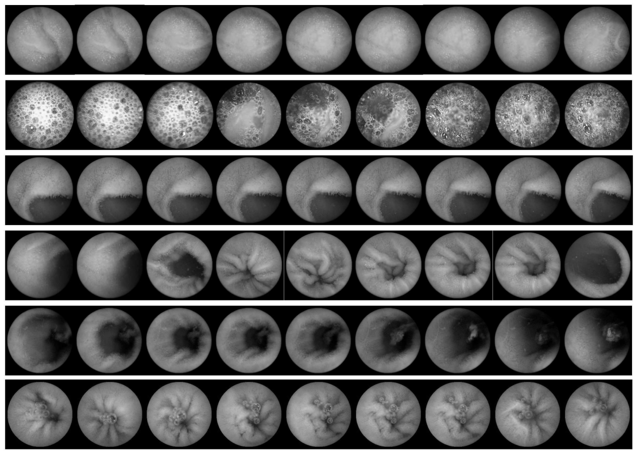

Figure 2: Some different intestinal video images. From top to bottom: Static Sequence, Turbid frames, Tunnel Sequence, Occlusive contraction, Non-Occlusive contraction and Sustained Contraction.

Once the important information at frame level is extracted the set of video features are defined to characterize symptom of intestinal disease. These features are the following: 1. Percentage of tunnel in valid frames, 2. percentage of tunnel in all video, 3. mean of static value of tunnel frames, 4. mean length of tunnel sequences with gaps ≤ 10 frames, 5. median of the length of tunnel sequences with gaps ≤ 10 frames, 6. mean length of tunnel sequences with gaps ≤ 60 frames, 7. median of the length of tunnel sequences with gaps ≤ 60 frames, 8. turbid percentage in all video, 9. percentage of turbid static, 10. mean of static value of turbid frames, 11. mean length of turbid sequences, 12. percentage of wrinkle frames in valid frames, 13. percentage of wrinkle frames in all video, 14. mean of wrinkle level in all video, 15. mean length of wrinkle sequences, 16. percentage of frames with low value of wrinkles, 17. percentage of frames with high values of wrinkles, 18. median of the length of wrinkle sequences, 19. percentage of static frames in valid frames, 20. percentage of static frames in all video, 21. mean length of static sequences, 22. mean of static value of valid frames, 23. median of the length of static sequences, 24. number of contractions per minute in valid frames, 25. number of contractions per minute in all video,

 ISBN 978-950-605-505-9

26. percentage of contractions with wrinkles, 27. percentage of non-occlusive contractions, 28. number of sustained contractions per minute in all video.

The binary classification problem takes as input a N-dimensional training data set $D = \{(\mathbf{x}^n, y_n)\}_{n=1,..N}$, where $\mathbf{x}^i \in \mathbb{R}^d$ is a sample and y_i is an integer value representing its corresponding label. The two classes are formed by the video samples with and without severe intestinal motility dysfunctions and are denoted C_1 and C_2 respectively. The dimension d of the input space is initially the number of features firstly defined by the experts.

It is known that the result accuracy of a classifier can be improved by performing an appropriate feature selection and determining which the most suitable subset of variables to consider during the classification process is. For this purpose, we apply the method presented in [4]. This method uses a Laplace prior to promote sparsity on the Relevance Vector Machine classifier parameters, as well as on the distribution of the selected features. Formally, given the samples of a 2-class classification problem, the negated log-likelihood estimator for the parameters set $\mathbf{w} = (w_0, ..., w_N)$ is defined as follows:

$$L(D, \mathbf{w}) = -\sum_{i=1}^{N} P(y = 1|x_i)^{y_n} P(y = 0|x_i)^{1-y_n}. \quad (1)$$

To promote sparsity on the selected vectors the loss function $L(D, \mathbf{w})$ is penalized through a constraint such as $R(\mathbf{w}) = \|\mathbf{w}\|_1$, getting the following loss function $G(\mathbf{w}) = L(D, \mathbf{w}) + \lambda R(\mathbf{w})$. The estimate for $\mathbf{w}$ when using this model can be interpreted as a Bayesian posterior mode estimate when the prior on the parameters are independent double-exponential (Laplace) distributions [7]. This is a simple formulation of the RVM, that can be called the Laplacean RVM (**L-RVM**).

To control the activation of the features, a new parameters vector $\mathbf{v} = (v_1, ..., v_d)$ is considered. Then, the global loss function for the RVM is constrained by an extended parameter set $\Omega = (\mathbf{w}, \mathbf{v})$:

$$\Gamma(\Omega) = L(D, \mathbf{w}, \mathbf{v}) + \lambda Q(\mathbf{w}), \quad (2)$$

and $Q(\mathbf{w}) = \|\mathbf{w}\|_1 + \|\sigma_k^d(\mathbf{v})\|_1$, where $\sigma_k^d : \mathbb{R}^d \to \mathbb{R}^d$ is defined as $\sigma_k^d(a) = (\sigma_k(a_1), ..., \sigma_k(a_d)), \forall a = (a_1, ..., a_d) \in \mathbb{R}^d$, k any positive real value, and $\sigma_k : \mathbb{R} \to \mathbb{R}$ is the sigmoid function defined as $\sigma_k(a) = \frac{1}{1+\exp(-ka)}, \forall a \in \mathbb{R}$.

The loss function (2) represents a preference for solutions that uses a small set of components from a small set of samples. The optimization of this functional is carried out using the Boosted Lasso algorithm proposed in [4].

Here, we tackle with an imbalanced problem. It is difficult to find subjects with the severe motility dysfunctions that we studied (elements of the class C_1), whereas many healthy subjects are available (elements of the class C_2). Moreover, in our field of work, it is easy to dispose of another set of subjects without a clear diagnostic. This data set is called *unsupervised* data set and is denoted X_u. The *supervised* data set is formed by the elements of the two classes and denoted $X_s = C_1 \cup C_2$. Standard classifiers only use supervised data. However, we propose to use a semi-supervised learning classifier in order to profit from the unsupervised data set. In particular, we apply an approach called self-training algorithm which claims to increase the too small number of data samples using the new data. The algorithm works as follows: create an initial training set with the supervised data set, classify the unsupervised samples, add the example classified with the higher score as patient. Iterate until any of the new examples are classified as patient. Finally, all the examples that are not classified as patient are included in the healthy subject set.

3 Experimental Results

We considered a set of videos obtained using the wireless endoscopic capsule developed and provided by Given Imaging, Ltd., Israel [8]. All of them were generated at same conditions, patients and healthy volunteers were in fasting (without eating and drinking in the previous 12 hours to the study), at Digestive Diseases Department of General Hospital ”Vall d’Hebron” in Barcelona, Spain.

Our experimental data set was formed by 65 videos: 15 from class C_1 and 50 from class C_2. Healthy volunteers were randomly selected from a big pool of subjects without symptoms and with a very low probability to be patients. In order to alleviate the problem of the limited size we considered a set of unsupervised data formed by 17 videos. The final training set X was formed by 82 samples in total, $X = X_s + X_u$. The unsupervised data were included using the self-training algorithm. It labelled 11 of the 17 unsupervised examples as patients and the remaining 6 as healthy subjects.

We performed two tests using a leave-one-out validation method [9] over the data set. The first test used the original RVM classifier applied with the set of 28 features enumerated before. The second test used the JFSCL for feature selection and classification. The results were validated using several measures that are described in terms of true positives (TP), true negatives (TN), false positives (FP) and false negatives (FN) as follows: *Error*= $FP + FN$, *Sensitivity*, *Specificity*, *Precision* and *False Alarm Ratio*. The unsupervised data are only used as training, not to test.

In both tests we made 3 errors in the set X_s (65 subjects): 1 FN (one patient was considered as healthy subject) and 2 FPs (two healthy subjects were classified as patients). The Table 1 summarizes the obtained results, the first row corresponds to the results applying only RVM and the second one are the ones of JFSCL method. The last column

Table 1: Classification Results

	Error	Sens.	Spec.	Prec.	FAR	m
JFSCL	4.61%	93.33%	96.00%	87.50%	13.33%	14.87
RVM	4.61%	93.33%	96.00%	87.50%	13.33%	28

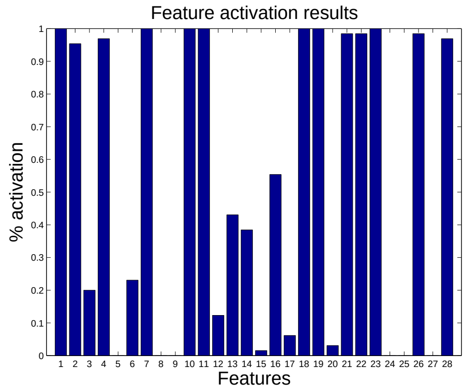

Figure 3: Feature activation results

displays m, the mean of the number of features used in the tests of the cross validation, $m = \#\text{features}/N$. In this case the total number of tests is $N = 65$.

Figure 3 shows the distribution of the activations for each feature in the leave-one-out validation experiment. As it can be seen the activation is quite robust choosing the same features in most of the tests. We considered the features disabled in less than a 10% of the tests as non relevant. These were the nine following features: 2, 8, 11, 12, 14, 16, 21, 26, 27. The features which were activated in more than a 80% of the test were considered as the more relevant ones. These were the next thirteen: 1, 4, 5, 6, 15, 17, 18, 19, 20, 22, 24, 25, 28.

4 Conclusion

In this paper we propose a method based on a joint feature selection and classifier training for diagnosis of severe intestinal motility dysfunctions. The method is applied to data obtained with the new Wireless Capsule Video Endoscopy, exploiting visual information to characterize intestinal motility dysfunctions. The diagnosis results are comparable to the ones obtained using the most applied diagnosis test, the manometry. Therefore, our contribution represents an important step towards the introduction of the WCVE for intestinal motility diseases diagnosis. The main advantage of this diagnosis test, compared with the current one, is that it is non-invasive and does not need monitorization. The feature selection method is able to procure the most relevant features among the ones defined by the experts, for the classification, reducing significatively the dimensionality and maintaining the accuracy.

Finally, the variability of symptoms of this kind of diseases brings us to consider, as future work, the problem of characterization of different groups of patients.

Acknowledgements

This work was supported in part by a research grant from Given Imaging Ltd., Yoqneam Israel, and Hospital Universitari ”Vall d’Hebron” Barcelona, Spain, as well as the project TIN2006-15308-C02.

References

[1] Guyon, I., Elisseeff, A.: An introduction to variable and feature selection. Journal of Machine Learning Research **3** (2003) 1157–1182

[2] Weston, J., Mukherjee, S., Chapelle, O., Pontil, M., Poggio, T., Vapnik, V.: Feature selection for SVMs. In: NIPS, MIT Press (2000) 668–674

[3] Krishnapuram, B., Hartemink, A.J., Carin, L., Figueiredo, M.: A bayesian approach to joint feature selection and classifier design. IEEE Trans. Pattern Anal. Mach. Intell **26**(9) (2004) 1105–1111

[4] Lapedriza, A., Seguí, S., Masip, D., Vitrià, J.: A sparse bayesian approach for joint feature selection and classifier learning. Technical Report CVC (2007)

[5] Hansen, M.B.: Small intestinal manometry. Physiological Research **51** (2002) 541–556

[6] Zheng, M.M., Krishnan, S.M., Tjoa, P.: A fusion-based clinical support for disease diagnosis from endoscopic images. Computers in Biology and Medicine **35**(3) (2005) 259–274

[7] Effron, B., Hastie, T., Johnstone, I., Tibshinrani, R.: Regression shrinkage and selection via the lasso. J. Royal. Statist. Soc. B. **58(1)** (1996) 267–288

[8] Given Imaging, L.: (http://www.givenimaging.com)

[9] Stone, M.: Cross-validatory choice and assessment of statistical predictions (with discussion). J. Royal Statistical Soc. B **36** (1974) 111–147