ESANN 2015 proceedings, European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning. Bruges (Belgium), 22-24 April 2015, i6doc.com publ., ISBN 978-287587014-8. Available from http://www.i6doc.com/en/.

On the use of machine learning techniques for the analysis of spontaneous reactions in automated hearing assessment

Verónica Bolón-Canedo, Alba Fernández, Amparo Alonso-Betanzos, Marcos Ortega and Manuel G. Penedo *

Department of Computer Science - University of A Coruña
Campus de Elviña s/n 15071 - A Coruña, Spain

Abstract. Lack of hearing is one of the most frequent sensory deficits among elder population. Its correct assessment becomes complicated for audiologists when there are severe difficulties in the communication with the patient. Trying to facilitate this task, this paper proposes a methodology for the correct classification of eye gestural reactions to the auditory stimuli by using machine learning approaches. After extracting the features from the existing videos, we applied several classifiers and managed to improve the detection of the most important classes through the use of oversampling techniques in a novel way. This methodology showed promising results, with true positive rates over 0.96 for the critical classes and global classification rates over 97%, paving the way to its inclusion in a fully automated tool.

1 Introduction

Hearing loss is the third most prevalent chronic health condition facing older adults [1]. The standard test for the clinical evaluation of hearing loss is the audiometry [2], a behavioral test where the hearing thresholds of the patient are evaluated in order to diagnose his or her hearing capacity. Since this evaluation is a behavioral test, it requires a high interaction and understanding between patient and audiologist. This need of communication is what causes serious difficulties when the patients suffer from cognitive decline or other communication difficulties. A typical interaction is not possible with this particular group of patients, instead, audiologists argue that there are some unconscious and spontaneous reactions that may correspond with involuntary signs of perception. These spontaneous reactions are gestural reactions that, in most cases, are expressed in the eye region. Changes in the gaze direction or an exaggerated eye opening might be interpreted as signs of perception to the auditory stimulus. The detection and interpretation of these gestural reactions requires broad experience from the audiologist, a high degree of concentration and it is very prone to errors.

All these circumstances highlight the need for an automatic tool which supports the audiologist in the evaluation of patients with cognitive decline. This method must be focused on the eye region in order to detect eye gestural reactions to the sound. To that end, in this proposal we are going to analyze the eye movements using optical flow information and color information from the sclera, the white area in the eye. The

*This research has been supported in part by the Consellería de Industria (research project GRC2014/035) and Consellería de Cultura, Educación e Ordenación Universitaria (through the agreement for the Singular Research Center CITIC) of the Xunta de Galicia and also by the Secretaría de Estado de Investigación of the Spanish Government through research projects TIN2011-25476 and TIN2012-37954.

ESANN 2015 proceedings, European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning. Bruges (Belgium), 22-24 April 2015, i6doc.com publ., ISBN 978-287587014-8. Available from http://www.i6doc.com/en/.

justification for using these features and an analysis of their behavior has been studied in detail in [3] and [4]. These two sources of information must be identified and characterized in order to determine if a movement has occurred as a reaction to the auditory stimulus. In order to determine if a significant movement has occurred, four eye movement categories were established in accordance with the audiologist. These four categories are: eye closure, eye opening, gaze shift to the left and gaze shift to the right. Two of these categories are typically associated with gestural reactions to the sound: gaze shift to the left and gaze shift to the right. This is motivated because patients with cognitive decline or severe communication difficulties tend to stay still and passive during the hearing evaluation, but, when they perceive an auditory stimulus they usually tend to direct their gaze to the side on which they perceive the sound. However, not every patient shows the same reaction, so the detection of the other eye movements could be relevant in particular cases.

After extracting the final features from the videos, machine learning techniques are applied aiming at automatically classifying the data into one of these four classes. Specifically, we apply several classifiers and then try to improve the true positive rate of the most important classes by using oversampling techniques. It is expected that the proposed methodology will enable the proper assessment of patients when no interaction is possible with high classification accuracy rates.

2 Methodology

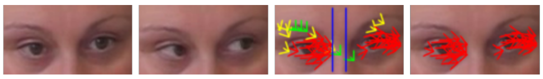
As depicted in the Introduction, the development of an automatic solution capable of analyzing the eye movements and detecting gestural reactions to the auditory stimuli would be very helpful in the hearing assessment of patients with cognitive decline. This automatic solution will receive as input a video sequence recorded during the development of the hearing assessment, and it is going to be analyzed frame by frame.

2.1 Location of the Region of Interest (ROI)

The first step of the methodology is the location of the region of interest, which in this case corresponds with the eye region. After that, we use two sources of information: the optical flow and the white color distribution of the sclera.

2.2 Optical flow information

In the first part, the motion is estimated by applying the iterative Lucas-Kanade [5] optical flow method with pyramids [6] over the ROI. Optical flow has shown optimal results in the identification of general and unconstrained movements produced by expression changes. After some corrections, the obtained vectors from the optical flow are categorized according to their strength since only the strong vectors are considered as significant movements. A sample of the obtained vectors and their classification can be observed in Fig. 1. In order to be able of reliably distinguishing the patient's movements it is necessary to characterize the movement using as base a set of properties associated with that movement. The considered properties are: orientation, magnitude and dispersion. The vector orientation provides information about the direction of the movement. For the definition of these descriptors, vectors are divided in eight different equally distributed ranges according to their angle and the first eight vectors of the descriptor correspond to the number of vectors in each of these ranges. The next eight values are associated to the vector's magnitude. With the vectors grouped by ranges, the average of the module of the vectors is calculated. These features provides information about the intensity of the movement, allowing to distinguish between strong and soft movements. Finally, the dispersion contributes with other eight values to the descriptor. The computation of the dispersion is also considered by range, according to the angle of the vectors. This methodology is further detailed in [3].



(a) (b) (c) (d)

Fig. 1: Sample optical flow images: Optical flow is computed between (a) and (b). Optical flow results in (c): green vectors for softer movements, yellow for intermediate and red for strongest movements. Stronger vectors in (d) after filtering the rest.

2.3 Color information from the sclera

In order to detect the direction of the gaze, we use information about the white color distribution of the sclera, so it is necessary to accurately delimit the eye's boundaries. The first step is the pupil location, which serves as a reference for the rest of the process. A combination of interest operators and edge information is used to located the eye's boundaries. In order to determine the direction of the gaze, we need to estimate the amount of white in the eye, using as reference the characteristic points previously obtained (pupil's center and eye's boundaries). For the characterization of the movement, we are going to compute a gray level distribution representing the gray level for each one of the pixels located in the line connecting both eye's corners. The result of this step can be observed in Fig. 2 (left). The gray level distribution can be divided into three areas of interest, i.e: iris, left side of the sclera and right side of the sclera. This way, starting from the pupil's center we go through the gray level distribution, both to the right and to the left, until we detect the first white pixel that indicates the boundary between the iris and the sclera. As a result of this step, the distribution of the delimitation of these three areas can be observed in Fig. 2 (right). For the feature vector, we include 30 values for the gray level distribution (the line connecting both corners is normalized to 30 values) and 3 extra values which represent the total values for: the left side of the sclera, the right side of the sclera, and for the left and right sides together. Furthermore, in order to represent changes in the gaze direction, we compute the difference between the gray level distribution of the actual frame *i* with the gray level distribution of the frame *i-3* (justified in [3]), which implies a contribution with 33 more values.

2.4 Classification

Finally, the descriptor of a ROI has to be classified into one of the four possible categories. For this task, several well-known classifiers available at the widely-used Weka

ESANN 2015 proceedings, European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning. Bruges (Belgium), 22-24 April 2015, i6doc.com publ., ISBN 978-287587014-8. Available from http://www.i6doc.com/en/.

tool [7] were employed. A 10-fold cross validation is performed in order to analyze the efficiency of the different methods.

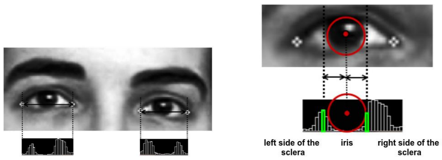


Fig. 2: On the left one sample gray level distribution, on the right the delimitation of the three areas of interest over the gray level distribution.

3 Experimental results

The proposed methodology has been tested on a normalized dataset which is composed of 1180 samples and 90 features. Four different classes are included: eye closure (408 samples), eye opening (310 samples), gaze shift to the left (230 samples) and gaze shift to the right (232 samples). Notice that the most important classes to detect are gaze shift to the left and to the right, since they are associated with gestural reactions to the sound. However, not every patient shows the same reaction, so the detection of the other eye movements could be relevant in particular cases. Table 1 shows the classification results using different algorithms, trying to find the most appropriate one for this problem.

Table 1: Classification results using a 10-fold cross validation for different algorithms

			True positive rate			
Method	Accuracy	ROC Area	Closure	Opening	Left	Right
C4.5	0.9000	0.9343	0.9152	0.9083	0.8948	0.8666
Naive Bayes	0.9331	0.9848	0.9487	0.9559	0.9447	0.8713
k-NN	0.9593	0.9721	0.9747	0.9532	0.9619	0.9368
SVM	**0.9712**	0.9899	**0.9838**	**0.9623**	**0.9787**	**0.9579**
RandomForest	0.9585	**0.9972**	0.9714	0.9567	0.9666	0.9357

In light of the results depicted in Table 1, the highest accuracy and TP rates were obtained with SVM. This fact is not surprising at all, since Support Vector Machines have demonstrated to be successful modeling and prediction tools for a variety of applications such as DNA microarray classification [8] or tear film lipid layer classification [9]. Regarding the true positive rates, it is easy to note that these algorithms are often biased towards learning the majority class [10], leading to higher misclassification rates for the minority class instances (gaze left and gaze right). Furthermore, it is worth pointing out that the minority classes are the ones that have the highest interest in the problem at hand, so we will try to improve their classification.

Numerous techniques are used to deal with imbalanced datasets in classification, among them we would like to highlight the use of *oversampling* methods, which aim to balance the class distribution by adding to the new dataset instances from the minority class. In particular, the SMOTE algorithm [11] (available in Weka) is an oversampling method that adds synthetic minority class examples to the original dataset until the

ESANN 2015 proceedings, European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning. Bruges (Belgium), 22-24 April 2015, i6doc.com publ., ISBN 978-287587014-8. Available from http://www.i6doc.com/en/.

class distribution becomes balanced. In order to do so, the SMOTE algorithm generates the synthetic minority class examples using the original minority class examples in the following way: the SMOTE algorithm searches the k nearest neighbors of the minority class sample that is going to be used as base for the new synthetic sample. Then, in the segment that unites the minority class sample with one or all of its neighbors, a synthetic sample is randomly taken and is added to the new oversampled dataset.

As stated before, the usual approach is to replicate the minority classes until the class distribution becomes balanced, so we applied an oversampling rate of 100% on the classes gaze left and gaze right. In the second row of Table 2 we can see the results of this experiment, in which the true positive rates for the two classes that are the most important in this problem (gaze left and gaze right) have increased, while maintaining the global accuracy. Based on some ideas raised in [12], and considering that what was of real interest for us was to improve the identification of these two minority classes, we performed some experiments increasing the oversampling rate to 200%, 300% and 400%, as can bee seen in the remaining rows of Table 2. For the sake of comparison, the first row shows the classification results when no oversampling technique is applied.

The best results for the gaze shifts were obtained when the level of oversampling is 400%. In this case, we achieve maximum true positive rates for the detection of the most important classes (gaze left and gaze right) whilst the global accuracy has been deteriorated in less than 0.3%. However, with 300% oversampling, we can maintain the global accuracy with an important increase in the detection of gaze shifts. Notice that, with these configurations, the TP rates for the detection of closure and opening decrease, but they are still in acceptable values considering their lower incidence in the detection of gestural reactions.

Table 2: SVM classification results using a 10-fold cross validation and applying different levels of oversampling

			True positive rate			
% Oversampling	Accuracy	ROC Area	Closure	Opening	Left	Right
0	**0.9712**	0.9899	**0.9838**	0.9623	0.9787	0.9579
100	0.9703	0.9907	0.9793	0.9589	0.9829	0.9607
200	0.9695	0.9898	0.9773	**0.9625**	0.9787	0.9615
300	**0.9712**	0.9901	0.9769	0.9589	0.9860	0.9667
400	0.9686	**0.9909**	0.9723	0.9528	**0.9882**	**0.9707**

4 Discussion and conclusions

In this paper we propose a methodology for the correct classification of eye gestural reactions to the auditory stimuli in order to facilitate the hearing assessment of patients when no cooperation exists. There are four eye movement categories: eye closure, eye opening, gaze shift to the left and gaze shift to the right. The two gaze shifts are typically associated with gestural reactions to the sound, so its correct detection is of utmost importance. However, these two categories are the ones with the smallest number of classes, so we proposed to include oversampling techniques in our methodology.

After extracting the features from the existing videos, we applied several state-of-the-art classification algorithms to determine which one was more appropriate. Since

SVM clearly outperformed the other techniques, we decided to use it for our case study. Then, and trying to increase the true positive rates for the two classes of interest (gaze left and gaze right), we opted for applying oversampling techniques, in particular, SMOTE. The novelty of this work lies in the fact that, instead of replicating the minority classes until they are balanced with respect to the majority classes, we decided to apply higher oversampling rates so that the minority classes have now much more instances than the previous majority classes. In this way, our hypothesis was that by forcing the classifier to learn from a data where the two classes of interest were the majority classes, their true positive rates would increase. Indeed, by choosing an oversampling rate of 300%, we were able to obtain true positive rates over 0.986 and 0.966 with a global classification accuracy over 97%.

In conclusion, the proposed methodology can classify the different movements with reasonable detection rates, especially for the most important classes. This methodology has shown encouraging and positive results, paving the way to its inclusion in an fully automated tool. As future work, there is a need of more video sequences of patients with this particular conditions so a more comprehensive analysis can be conducted. Also, we plan to try feature selection methods to check if accuracy can be improved.

References

[1] J.G. Collins. *Prevalence of Selected Chronic Conditions: United States 1990-1992*, volume 10 (194). Hyattsville, MD: National Center for Health Statistics, Public Health Service, National Center for Health Stadistics, 1997.

[2] A.C. Davis. The prevalence of hearing impairment and reported hearing disability among adults in Great Britain. *Int J Epidemiol.*, 18:911–917, Dec. 1989.

[3] A. Fernández, M. Ortega, M. G. Penedo, C. Vazquez, and L. Gigirey. A methodology for the analysis of spontaneous reactions in automated hearing assessment. *IEEE Journal of Biomedical and Health Informatics*, 2014.

[4] A. Fernández, J. de Moura, M. Ortega, and M. G. Penedo. Detection and characterization of the sclera: evaluation of eye gestural reactions to auditory stimuli. In *10th International Conference on Computer Vision Theory and Applications (VISAPP), submitted*, 2015.

[5] B. D. Lucas and T. Kanade. An iterative image registration technique with an application to stereo vision. In *Proceedings of the 7th international joint conference on Artificial intelligence - Volume 2*, IJCAI'81, pages 674–679, San Francisco, CA, USA, 1981. Morgan Kaufmann Publishers Inc.

[6] J.Y. Bouguet. Pyramidal implementation of the Lucas-Kanade feature tracker: Description of the algorithm. *Intel Corporation, Microprocessor Research Labs*, 2000.

[7] M. Hall, E. Frank, G. Holmes, B. Pfahringer, P. Reutemann, and I. H. Witten. The weka data mining software: an update. *ACM SIGKDD Explorations Newsletter*, 11(1):10–18, 2009.

[8] F. F. G. Navarro. *Feature selection in cancer research: microarray gene expression and in vivo 1h-mrs domains*. PhD thesis, Universitat Politècnica de Catalunya, 2011.

[9] B. Remeseiro, M. Penas, A. Mosquera, J. Novo, M.G. Penedo, and E. Yebra-Pimentel. Statistical comparison of classifiers applied to the interferential tear film lipid layer automatic classification. *Computational and Mathematical Methods in Medicine*, 2012.

[10] V. López, A. Fernández, Sa. García, Va. Palade, and F. Herrera. An insight into classification with imbalanced data: Empirical results and current trends on using data intrinsic characteristics. *Information Sciences*, 250:113–141, 2013.

[11] N. V. Chawla, K. W. Bowyer, L. O. Hall, and W. P. Kegelmeyer. Smote: synthetic minority oversampling technique. *arXiv preprint arXiv:1106.1813*, 2011.

[12] S. del Río, V. López, J. M. Benítez, and F. Herrera. On the use of mapreduce for imbalanced big data using random forest. *Information Sciences*, 285:112 – 137, 2014.