MULTI-CLASS CASE-BASED REASONING FOR MEDICAL APPLICATIONS: AN EXPLORATORY STUDY

M. Ortega-Adarme [1] D. Viveros-Melo [2] X. Blanco Valencia [3] A. E. Castro-Ospina [4] S. Murillo Rendón [5] D. H. Peluffo-Ordónez [6]

[1,2]Universidad de Nariño [3]Universidad de Salamanca [4]Instituto Tecnológico Metropolitano [5]Universidad Autónoma de Manizales [6]Universidad Técnica del Norte

INTRODUCCTION

Case-based reasoning (CBR) is a problem solving approach that uses past experience to solve current problems. It may be viewed, as a perspective on human cognition, and as a methodology for building intelligent systems [1]. The core of the CBR is the case, usually indicates a problem situation; prior learning experience, which has been captured and can be reused to solve future problems. The life cycle for solving a problem using CBR is mainly carried out in four phases:

to identify the current problem and find a past case similar to the new case (retrieve), using the case and suggest a solution to the current problem (reuse/adaptation), evaluate the proposed solution (revise), and update the system to learn from experience (retain) [1]. The CBR has demonstrated to be an appropriate methodology for:

- Working with unstructured domains data or difficult knowledge acquisition situation, for example, many diseases are not well understood by formal models or universally applicable guidelines [2], [3].
- The training of health care professionals [3].
- When guidelines are available, they provide a general framework to guide clinicians, but require consequent background knowledge to become operational, which is precisely the kind of information recorded in practice cases; cases complement guidelines very well and help to interpret them [3].
- Domains where a deep model can represent the disease process, such as heart disease, often several diagnoses interact and the data are large, diverse and complex, difficult to be evaluated by traditional methods [4].

A study [3] concludes that some of the trends and opportunities that may be developed for CBR in the health science are oriented to the case representation where is necessary synthesizing adequate features for CBR, reducing the number of features in high dimensional data and one important focus will be how case-based reasoning can associate probabilities and statistics with its results by taking into account the concurrence of several ailments [5].

In order to adequately represent data and to avoid the inconveniences caused by its high dimensionality, we propose the use of variable selection and dimension reduction techniques in a preprocessing stage for CBR tasks, finally, we make a comparative study of multi-class classifiers to assess processed data performance.

CASE BASED REASONING

CBR is a problem solving approach that uses past experience to solve actual problems, by matching new problems with similar past situations. It may be viewed, as a perspective on human cognition, and as a methodology for building intelligent systems [6]. If the problem is solved, this new situation will be retained in order to solve the new ones. The CBR has been formalized as a four-steps process [7]: (1) RETRIEVE one or more of the most similar cases; (2) REUSE the solution of the retrieved case to solve the new problem; (3) REVISE the suggested solution; (4) RETAIN the experience of the new problem solution for future use.

The classical model of the problem solving cycle in CBR is shown in the Figure 1.

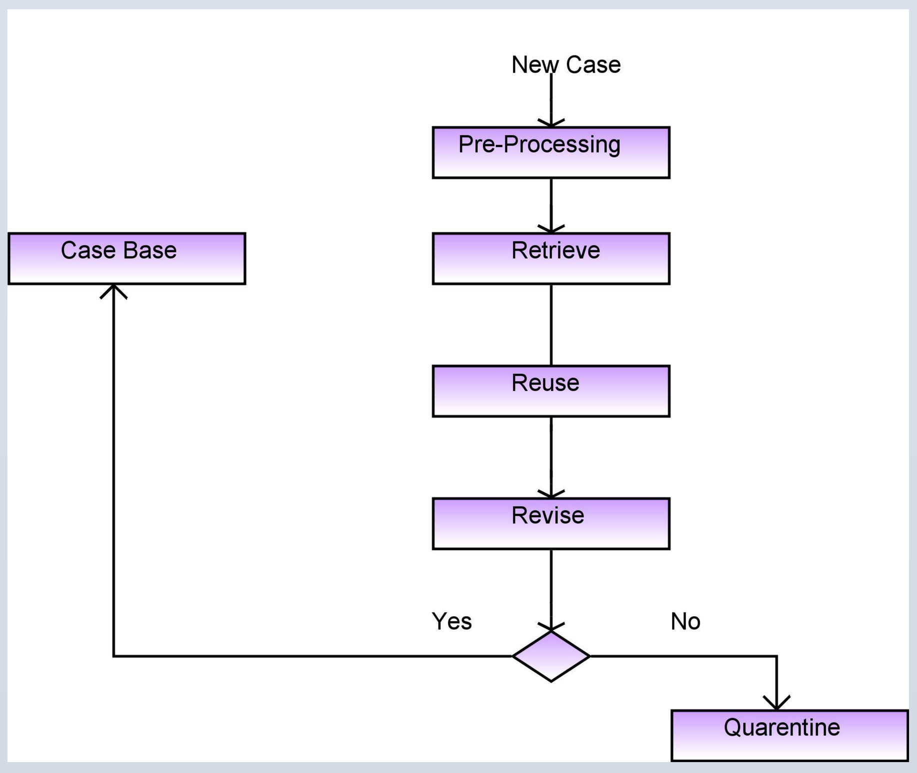

Figure 1. Cycle in CBR

EXPERIMENTAL SET-UP

For evaluating the proposed methodology, we used one databases from UCI Machine Learning Repository, named Cardiotocograms, contains 2126 fetal cardiotocograms belonging to different classes. This data set consists of 21 attributes. First, as preprocessing stage a variable selection procedure is employed. In this work, we use the socalled correlation based feature subset (CfsSubsetEval) algorithm, which evaluates the relevance of a subset of attributes by considering the individual predictive ability of each feature along with the degree of redundancy among them.

After performing variable selection and aiming to improve both visual inspection and classification performance, a dimensionality reduction stage is employed by using well known methods, namely Laplacian Eigenmaps (LE) and t-distributed stochastic neighbor embedding (t-SNE). As outcomes of the preprocessing stage, we obtain that Cardiotocograms database is reduced to 10 features. Subsequently, as part of the same stage, by using dimensionality reduction techniques Cardiotocogram database is reduced to a 2-, 3-, 5-, 8-dimensional space. For classification techniques, it should be stated out that a 20-fold cross-validation was performed to achieve unbiased results.

RESULTS

For Cardiotocograms dataset classes separability is evident in lower dimensions, i.e. 2D and 3D, as depicted in Figures 2(a) to 2(d) leading to outstanding results. For the Cardiotocograms database the best result was using the SVM classifier the error is 0.028 ± 0.016, improving the results obtained in [8] where they achieved an average accuracy of 0.9328.

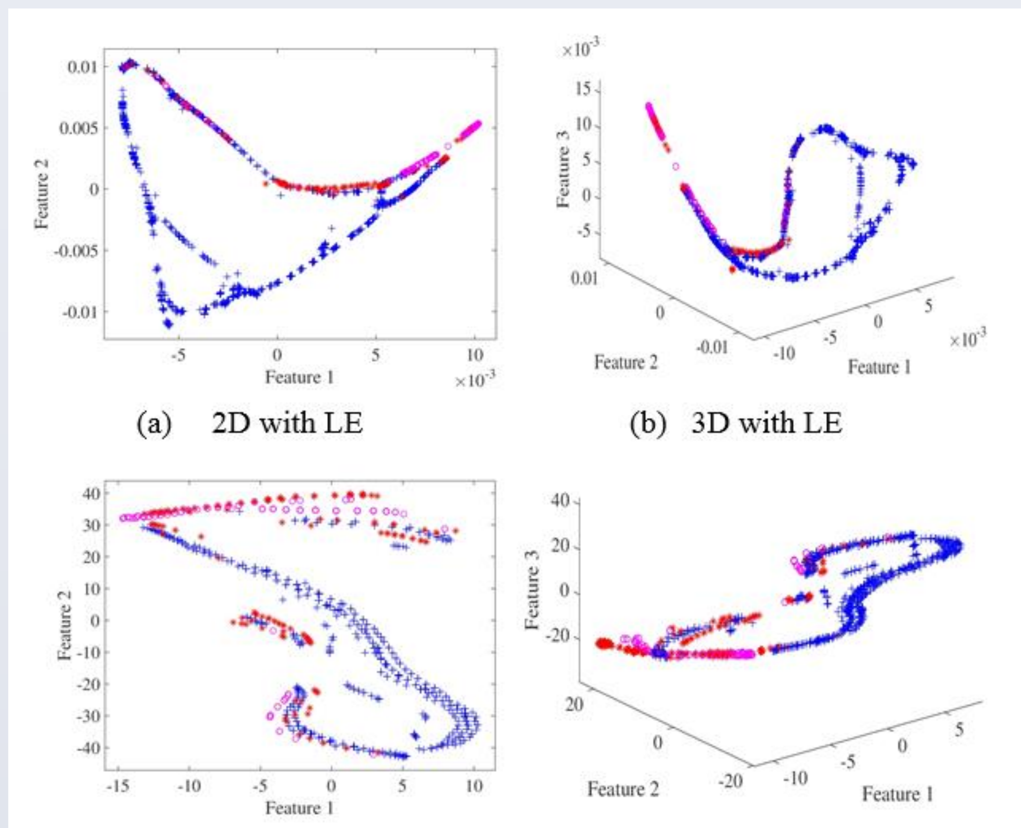

Figure 2. Low-dimensional scatterplots for Cardiotocograms database

REFERENCES

[1] J. Kolodner, Case-based Reasoning. San Francisco, CA, USA: Morgan Kaufmann Publishers Inc., 1993.

[2] J. M. Juárez Herrero, “Una aproximación multimodal al diagnóstico temporal mediante razonamiento basado en casos y razonamiento basado en modelos. aplicaciones en medicina,” Inteligencia Artificial. Revista Iberoamericana de Inteligencia Artificial, vol. 11, pp. 77–80, 2007.

[3] I. Bichindaritz, “Case-based reasoning in the health sciences: What’s next?” Artificial Intelligence in Medicine, vol. 36, no. 2, pp. 127–135, feb 2006.

[4] R. Vidya and G. Nasira, “Knowledge extraction in medical data mining: A case based reasoning for gynecological cancer - an expert diagnostic method,” ARPN Journal of Engineering and Applied Sciences, vol. 10, no. 9, pp. 3997–4001, 2015.

[5] M. Kwiatkowska and S. Atkins, “Case representation and retrieval in the diagnosis and treatment of obstructive sleep apnea: A semiofuzzy approach,” 2004.

[6] J. Kolodner. Case-based reasoning. Morgan Kaufmann, 2014.

[7] T. Selvy, V. Palanisamy, and S. Elakkiya. A novel approach for the prediction of epilepsy from 2d medical images using case based reasoning classification model. WSEAS Transactions on Computers, 12(6):243, 2013.

[8] Sundar.c, M.chitradevi, and G. Geetharamani, “Article: Classification of cardiotocogram data using neural network based machine learning technique,” International Journal of Computer Applications, vol. 47, no. 14, pp. 19–25, June 2012, full text available.

ACKNOWLEDGEMENTS

Authors would like to thank to the Facultad de Ingeniería en Ciencias Aplicadas as well as electronic engineering and telecommunications program from Universidad Técnica del Norte.