2014 IEEE International Conference on Fuzzy Systems (FUZZ-IEEE)
July 6-11, 2014, Beijing, China

Supervising classrooms comprising children with dyslexia and other learning problems with graphical exploratory analysis for fuzzy data: Presentation of the software tool and case study

Ana Maria Palacios
Department of Computer Science
and Artificial Intelligence
University of Granada, 18071 Granada, Spain
Email: palacios@decsai.ugr.es

Luciano Sanchez
Departamento de Informática
Universidad de Oviedo
33203 Gijon, Spain
Email: luciano@uniovi.es

***Abstract*—A case study is presented where a classroom comprising children with dyslexia and other learning problems is monitored with a proprietary software tool. This tool implements graphical exploratory analysis for fuzzy data, has classification and prediction capabilities, and outputs a compared visualization of the results of the tests applied to the children in the classroom and their expected evolution in the near future. The software also helps to detect different groups in the class, and could recommend professional treatment in case that some indicators were deemed abnormal. Data can be introduced without the help of a pshychologist thus visualization and diagnostic tasks can be performed by parents and teachers on their own.**

Keywords: Dyslexia; Low Quality Data; Data Mining; Visualization Algorithms.

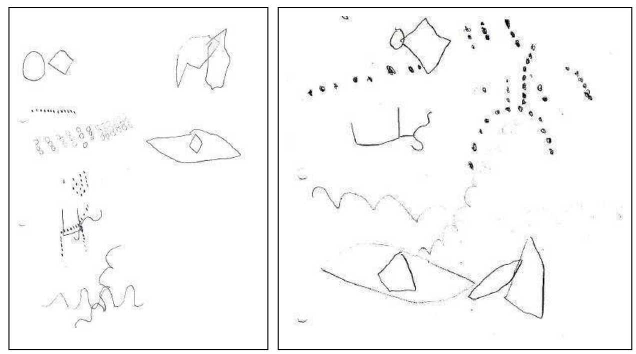

Fig. 1. Left: Bender's test of one child assigned to the class "no dyslexia". Right: The same test solved by a child for whom the expert could not decide between the classes "no dyslexia" and "control and revision".

I. INTRODUCTION

Dyslexia is a common learning disorder that manifests itself as a difficulty for the distinction and memorization of letters, lack of order and rhythm in the placement and poor structuring of sentences, affecting both reading and writing [9]. Dyslexia may also be defined as the learning difficulties of people whose IQ is normal and do not have physical or psychological problems that may explain these difficulties. The learning difficulties of a dyslexic child usually increase with time. Reading difficulties and bad understanding lead to poor school performance, low self esteem, and attitudes and behaviors that may affect the classroom. Because of this, early detection and reeducation of dyslexic children is crucial.

Different types of syndromes related to dyslexia exist, such as hyperactivity, attention deficit disorder or dysgraphia. They all share a certain degree of similarity, but differ in the acquisition of certain processes, such as identification, recognition or understanding. Failures or deficit in each of these processes produce a variety of different problems. Identifying the specific learning disorder is important, because each syndrome has different reeducation techniques and a different evolution in time. The diagnostic process involves both the teacher and the school phychologist. Notwithstanding the large amount of material, the work flow is currently done by hand, as explained below.

A. Work flow

Learning disorders are diagnosed by assessing the following factors:

1) The specific syndrome (dyslexia, hyperactivity, attention disorder, etc.)
2) The processes involved (recognition, understanding, etc.)
3) The acquisition level of each process

There are different tests for each process, and a large amount of information is collected for each student that must be analyzed before diagnosing (see Figure 1 for an actual sample of this kind of tests). This task is hard to solve by hand. However, up of our knowledge there are not software tools that help with this duty. The closest protocol might arguably be PRODILEX (from the Spanish *Protocolos de Detección y Actuación en Dislexia*) [10]. This protocol consists in a text document with a set of questions. The answers must be evaluated by the psychologist, who assigns a score to each child that measures the severity of his/her learning problems.

The school psychologist finds the learning disorders in the class and hands the results to the teacher, who must coordinate the classroom work, paying special attention to children with difficulties. Not only the syndrome but its evolution in time are taken into account. Analyzing the temporal evolution of children is also currently done by hand in Spanish schools, thus the need for a software tool that automates classroom supervision is clear.

B. Data mining and visualization techniques

The problem at hand does not require large databases but each individual is characterized by a high number of features, thus certain data mining algorithms are appropriate in this context.

Data mining encompasses different tasks [14], such as predicting or uncovering trends and relations hidden in the data. In those cases where the discovered knowledge is not easily deciphered, graphical tools can help the human expert to build his/her own cognitive model [2], [4] and to fully comprehend the output of the data mining process [3], [5], [12], [13].

The purpose of these techniques is to compress data comprising a high number of features into a low dimensional map, with a minimum loss of information. Related techniques are 2D and 3D scatterplots, matrixes of scatterplots, heat maps, survey plots, iconographic displays, dimensional stacking, polar charts, muti-dimensional scaling, Sammon plots, principal component and principal curve analysis, among others [3].

In this paper it is proposed that data mining-based visualization techniques may be used for analyzing the learning results in a classroom comprising children with dyslexia and other learning problems. Supervising the evolution of the children is also contemplated. By means of these techniques, supervisors can optimize their resources and detect the children with special needs earlier than it can be done with the standard protocols [9], [10].

In addition to this, in this problem there is a considerable fraction of missing data and vague features. These are originated in the subjective scoring of some of the tests. For instance, graphical tests do not comprise closed lists of items but drawings that are assessed with sentences like "lines are parallel" or "the shape is an square" (recall Figure 1). There a few works extending some of the aforementioned methods to vague data [11], however the graphical analysis of databases with missing or imprecise instances is an emerging research field. In this respect, in this paper it is proposed that the RadViz (Radial Coordinate visualization) method [1] can be used for analyzing and supervising classrooms comprising children with learning disorders such as dyslexia, if extended to fuzzy data.

Because of the mentioned reasons, in this work a software tool is proposed that presents the classroom information to psychologists, teachers and parents in an intuitive way. The organization of this paper is as follows: In Section II the software tool is introduced and its funcionality described. In Section III, technical details about the proposed visualization method are discussed. In Section IV a case study is shown. The paper concludes in Section V.

II. Description of the software tool

The software tool is designed to keep track of a database of evaluations of the childs in the classroom, and performs an analytic processing of this information that comprises diverse listings and graphical representations, whose technical details are given in the following section, as mentioned. An effort has been made to design an user-friendly tool that can be used by teachers and parents without computer training. Knowledge in dyslexia or learning problems is not assumed either.

Eleven categories of tests are managed:

- C0: Vocabulary
- C1: Verbal orders
- C2: Basic concepts
- C3: Reasoning
- C4: Visual-motor coordination
- C5: Verbal-additive memory
- C6: Perception of shapes
- C7: Spatial orientation
- C8: Auditive perception, rythm
- C9: Laterality
- C10: Pronunciation

The scores of the tests are in the format that the psychologist deemed to be more intuitive for each case: these can be numbers, intervals or fuzzy sets. In some cases, there are also missing evaluations that will be represented as an interval spanning the whole range of the variable. This is an improvement over the aforementioned PRODILEX, whose inputs are based on numbers and Likert scales.

As an example of this, in the second column of Table I the values assigned to three PRODILEX items are compared to the values introduced by teachers and parents in the proposed system for the same data. Observe that, in this particular case, conclusions drawn from a PRODILEX-based analysis will be based on the assumptions that there are not memory neither spatial orientation problems, but there are rhythm problems. However, the new system takes into account that "Memory" has a precise value of 7.5 out of 10, but the value of "Rhythm" is less specific: a range of values [2.9,4.2]. Lastly, three conflicting measurements $\{5.2, 6, 6.8\}$ were obtained in tests for "Spatial Orientation". The conclusions obtained with this richer data may be more informative. This will be discussed further in the case study.

The possible diagnostics of a child are subsets of the following list:

1) Not dyslexic
2) Control and revision (i.e., might be dyslexic but it is too soon to know, check again in the future)
3) Dyslexic
4) Other problem (learning disorder different than dyslexia, i.e., attention defficit, hyperactivity etc.)

The uncertainty in the diagnostic is also contemplated. Children may be assigned multiple labels from this list; for instance, a child may be labelled {Not dyslexic, Control and revision} showing that the psychologist doubted between these two alternatives.

Feature	Output of PRODILEX	Output of the proposed system
Memory , etc.	No	7.5
Rhythm	Yes	[2.9,4.2]
Spatial orientation	No	(5.2,6,6.8)

TABLE I. COMPARISON OF INFORMATIONS INTRODUCED TO PRODILEX AND TO THE PROPOSED SOFTWARE TOOL.

Finally, it is remarked that the new tool allows non-experts to find the most similar children in terms of the levels of the factors under study. This allows teachers and parents to know whether the evolution of the children is similar to that of other children with the same syndromes and seek for professional help if the differences are noticeable.

A. Functional description

The two main entry points in the application are Classroom overview and Classroom trends (see Figure 2). A short description of each follows:

1) *Classroom overview:* A visualization of a group of children, with the following options:
   a) Visualize a group of children: Any group of students are selected, and a map is generated that takes into account the desired factors (see Figure 3 for an example).
   b) Detect far from average children: The best/worse students are found with respect to the scores in the desired tests.
   c) Nearest students: A student is selected, and a list is produced comprising the students whose tests are similar to the selection.

   In all cases, a graphical visualization is shown, that is accompanied by the following listings (see Figure 4):
   - Information about the selected tests and factors.
   - Ranking from best to worst of the selected students, according to the chosen tests.
   - Explanation of the meaning of the symbols in the graphs, including the color code of the diagnosis made by the phsychologist.
2) Classroom trends: The evolution of a group of students is plotted. The source data for this analysis can be imported from the evaluations log file (see Figure 5), or the effect of a new learning technique applied to some students can be studied.

III. RADIAL COORDINATE VISUALIZATION OF HIGH-DIMENSIONAL FUZZY DATA

The visualization stage in the diagnostic tool is carried out by means of an extension of RadViz (Radial Coordinate visualization) [6] to imprecise data. Radviz is a visualization technique that maps a set of multivariate vectors onto a plane. Each point is held in place with springs that are attached at the other end to anchors (see Figure 6).

The strength of each spring depends on the value of the corresponding feature, and the spring forces are in equilibrium. If the number of processes is p, the anchors are at positions $(\cos(2k\pi/p), \sin(2k\pi/p))$, thus the value

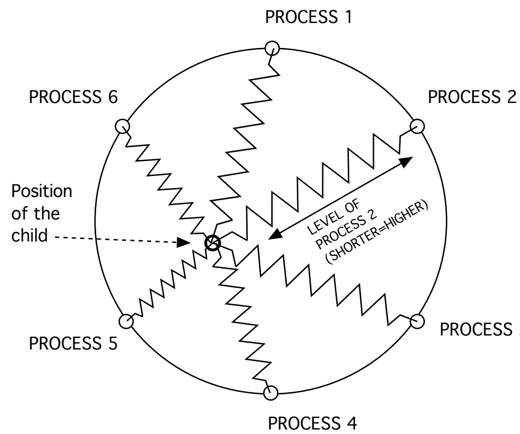

Fig. 6. The stiffness of each spring is proportional to the value of the feature. Similar children are projected into close map points.

$(v_1, v_2, \ldots, v_p)$ is mapped to the point

$$\left(\frac{\sum_{k=1}^{p} v_k \cos(2k\pi/p)}{\sum_{k=1}^{p} v_k}, \frac{\sum_{k=1}^{p} v_k \sin(2k\pi/p)}{\sum_{k=1}^{p} v_k} \right). \quad (1)$$

In these visualizations, individuals $\mathbf{v} = (v_1, \ldots, v_p)$ and $k \cdot \mathbf{v} = (k \cdot v_i, \ldots, k \cdot v_p)$ for $k > 0$ are mapped to the same point, and this is not desirable in this context because useful information is lost. This problem is solved by adding a new coordinate to each individual with the constant value 1, i.e. individuals have the form $(\mathbf{v}, 1) = (v_1, v_2, \ldots, v_p, 1)$. Interestingly enough, this decision introduces an instrumental anchor for the $p+1$-th process that can be given a practical use. Observe that the representations of $(\mathbf{v}, 1)$ and $(k \cdot \mathbf{v}, 1)$ are the same as that of

$$\left(\frac{v_1}{||\mathbf{v}||}, \ldots, \frac{v_p}{||\mathbf{v}||}, \frac{1}{||\mathbf{v}||} \right) \quad (2)$$

and

$$\left(\frac{v_1}{||\mathbf{v}||}, \ldots, \frac{v_p}{||\mathbf{v}||}, \frac{k}{||\mathbf{v}||} \right) \quad (3)$$

respectively, with $||\mathbf{v}|| = \sqrt{v_1^2 + \ldots + v_p^2}$. Generally speaking, this means that the distance between the map point and the $p+1$ anchor depends on the modulus of the individual. This added coordinate causes that an homogeneous percent gain in all processes is represented as a displacement in the $p+1$ spring, that would not be visible with the original RadViz visualization.

In addition to this, in this contribution it is also proposed that each child is represented by a vector of p fuzzy features, thus the vector being projected is

$$(\widetilde{V}_1, \widetilde{V}_2, \ldots, \widetilde{V}_p, 1) \quad (4)$$

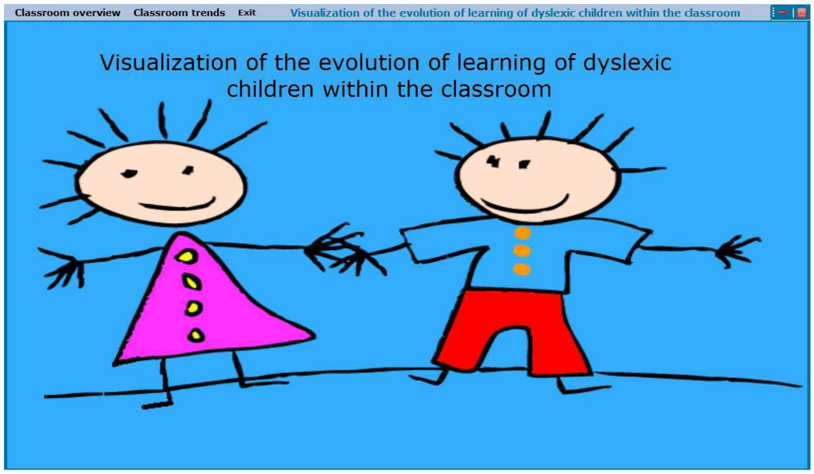

Fig. 2. Splash screen and entry points

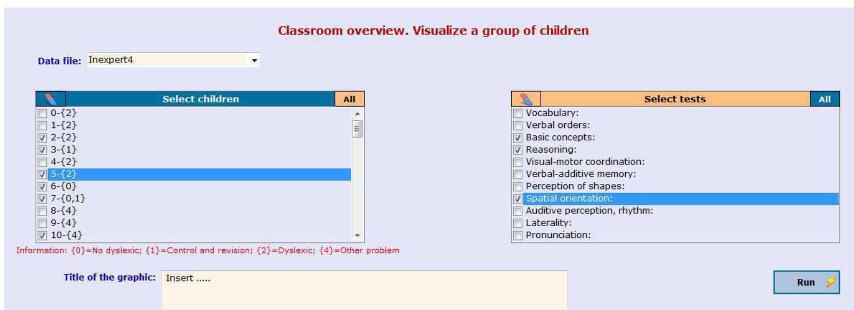

Fig. 3. Selection of a combination of students and tests

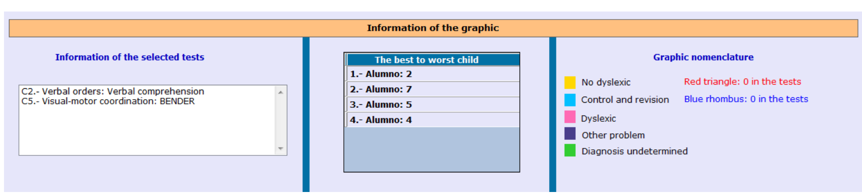

Fig. 4. Textual user information

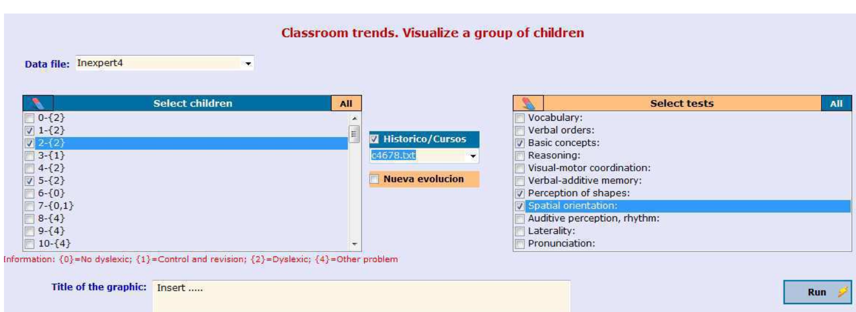

Fig. 5. Evolution of the learning, data taken from log file

which will be mapped to a fuzzy subset of the map by means of the extension principle [7]. The membership function of the map representation of a fuzzy individual is

$$\mu_{\mathrm{MAP}}(x_1,x_2) = \left\{ \max_{v\in\Re^{p+1}} \min_{k=1}^{p+1} \widetilde{V}_k(v_k) : \right.$$
$$x_1 = \frac{\sum_{k=1}^{p+1} v_k \cos(2k\pi/(p+1))}{\sum_{k=1}^{p+1} v_k},$$
$$x_2 = \frac{\sum_{k=1}^{p+1} v_k \sin(2k\pi/(p+1))}{\sum_{k=1}^{p+1} v_k},$$
$$\left. v = (v_1,\ldots,v_p,1) \right\}. \tag{5}$$

In the next section, the computer visualization of each of these fuzzy subsets of the map will be drawn by means of the ellipse that best fits their α-cut for a user-selected level.

IV. Case Study

In this section a case study comprising 65 psychological evaluations is described. Schoolchildren between 6 and 8 from Asturias (Spain) participated in this study. The supervision of this group is carried out with the proposed tool. Different analysis were performed and their results, consisting in several Fuzzy RadViz graphs, are presented and discussed. The color code used in all graphs is as follows:

- Pink: Dyslexic
- Yellow: Not dyslexic
- Dark blue: Learning problem different than dyslexia
- Light blue: Control and revision
- Green: Unknown diagnostic
- Blue diamond: Best marks for all tests
- Red triangle: Worst marks for all tests

A. Commented Results

The different views of the clasroom data are described now. Some of the analysis are focused in the detection of children with learning problems. In other cases, it was intended to refine doubtful diagnostics by means of a compared positioning of the individuals with unsure classification in the class map.

1) *Global visualization of the class as a whole group:* All children and all tests are simultaneously plotted in the same graph (see Figure 7). Observe that children labelled as "Not dyslexic" and "Control and revision" are near the diamond anchor, while "Dyslexic" and "Learning problem" are the furthest.
2) *Detect the presence of map areas related to the learning levels and problems:* The tool can be used to make a preliminary classification/screening of children without a psychological assessment, plotting them along the same graph as diagnosed children. It was found in Figure 8 that there is a reasonable separation between the map areas of "not dyslexic" and "control and revision"

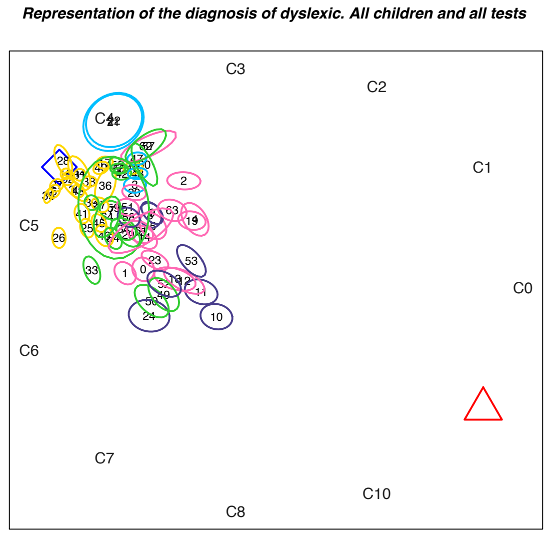

Fig. 7. Global visualization of the state of a whole classroom

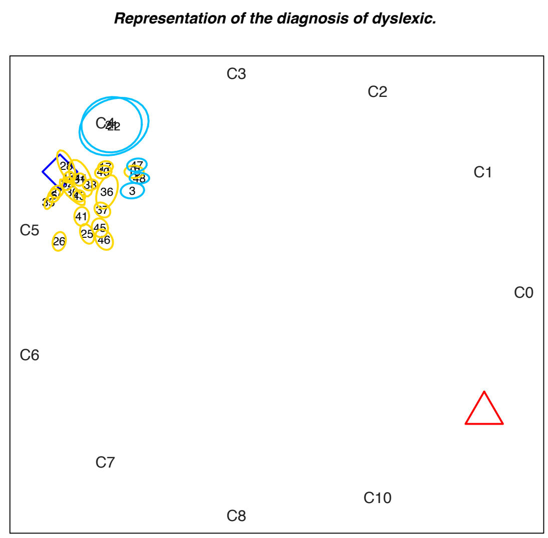

Fig. 8. Classification areas related to the learning levels and problems.

3) *Visualize the most appropriate class for a child with a doubtful diagnostic*: There are some children labelled "unknown diagnostic". In Figures 9, 10 and 11 it can be shown that children diagnosed as {Dyslexic, Unknown} are in the "Dyslexic" area, while children labelled as {Not dyslexic, Unknown} are not clearly in the "not dyslexic" area. However, children labelled {Other problem, Unknown} are far from children labelled "Other problem", thus their classification is not clear.
4) *Descriptive power of the different tests:* The effectiveness of the different tests can be assessed by plotting the same data with respecto to different subsets of indicators. In Figure 12 it is shown that the test C1 (Comprehension of orders) and C2 (basic concepts) are not deemed relevant for classifying dislexic children, because the separation between classes is lower than the separation achieved when all tests are considered.

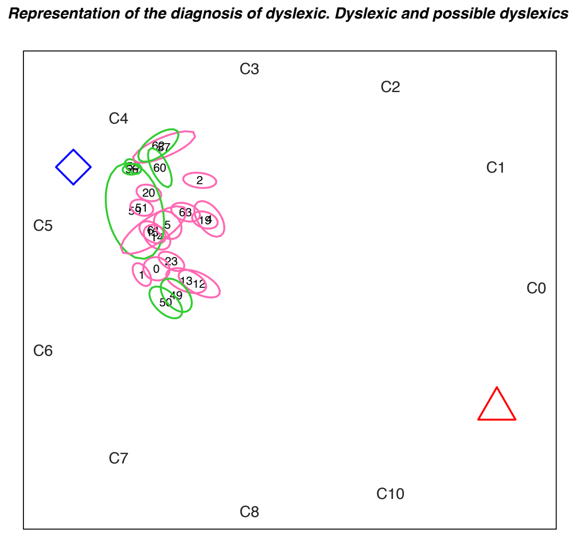

Fig. 9. Most appropriate class for a child with a doubtful diagnostic: Possible Dislexic

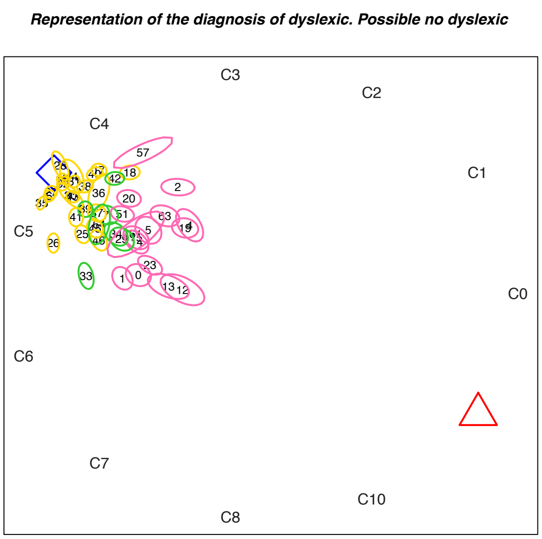

Fig. 10. Most appropriate class for a child with a doubtful diagnostic: Possible not dyslexic

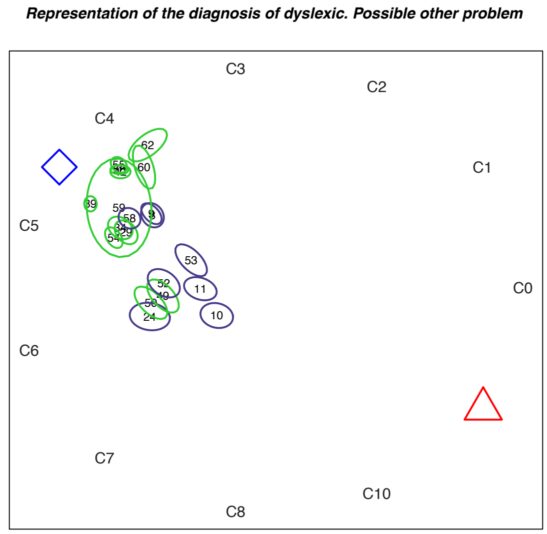

Fig. 11. Most appropriate class for a child with a doubtful diagnostic: Possible learning disorder

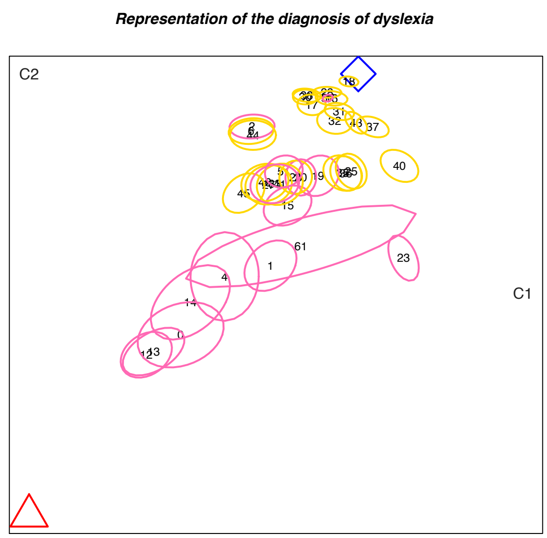

Fig. 12. Descriptive power of the different tests

5) *Fusion of classes:* In Figure 13 elements labelled "Control and revision" and "Dyslexic" were plotted together. In Figure 14, 'Control and revision" and "Not dyslexic". The density of the resulting cloud is higher in the first case, thus the fusion of "Control and revision" and "dyslexic" is preferred to the fusion of "Control and revision" and "Not dyslexic".
6) *Determine the children that are furthest from the average:* Given a set of tests, the children further from the average are probably in need of personalized attention. In Figure 15 the best 10 children are displayed with respect to the 11 tests. Observe that all of them are labelled as "Not dyslexic" or "Unknown", although these last ones are clearly in the "Not dyslexic" area. In Figure 16 the 10 worst are shown. All of them have learning problems or dyslexia but one. This individual should be re-examined by the psychologist.
7) *Position a new child in the classroom:* In Figure 17, four tests were selected. The position of a single child is compared to that of the other children in the class. Teacher or parents can determine the level of this child in the class by checking the labels of the nearest elements. In this case, the child probably has a learning problem because the nearest individuals have learning problems too. In particular, there is a high chance that the child is dyslexic because child number 23 is.
8) *Evolution of children:* The temporal evolution of the class can be tracked. Moreover, specific learning tasks may be assigned to children with learning problems. Comparing their capabilities before and after the task, the usefulness of the specific learning can be assessed. In this case (see Figure 18), a new task was programmed that is expected to improve results in

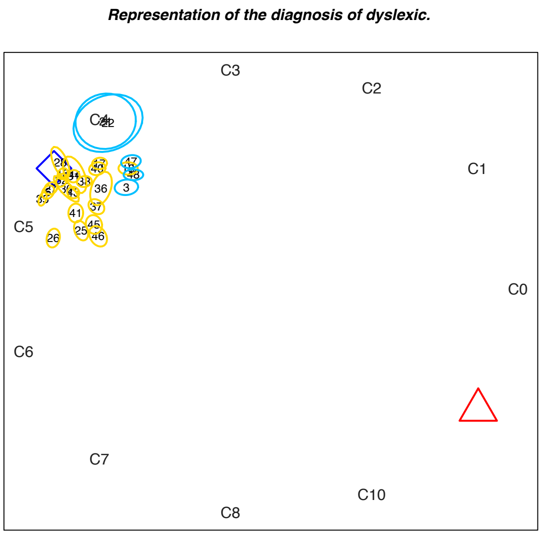

Fig. 13. Fusion of classes “Control and revision” and “dyslexic”

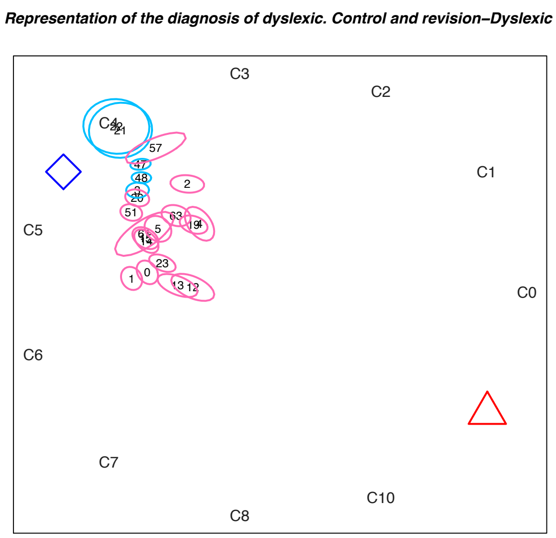

Fig. 14. Fusion of classes “Control and revision” and “not dyslexic”

tests C4, C6, C7 and C8, however the arrows show that the improvement is more significant for test C6 (shape perception), although the displacement is high not enough to alter the diagnostic of any child. On the contrary, a different learning task is assessed in Figure 19. A noticeable change is produced, mostly for dyslexic children, as shown by arrows pointing to the diamond anchor.

V. Concluding Remarks

A software tool is presented that allows parents and teachers to better comprehend the learning difficulties of a whole classroom. On the one hand, a graph showing similarities and differences between children for a customizable set of learning indicators allows segmenting the class into groups of similar children, optimizing the use of the material. On the other hand, evolution graphs are provided

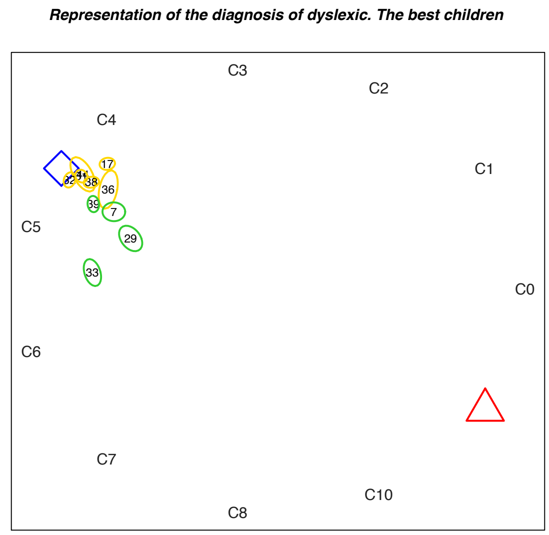

Fig. 15. Best 10 children with respect to all tests

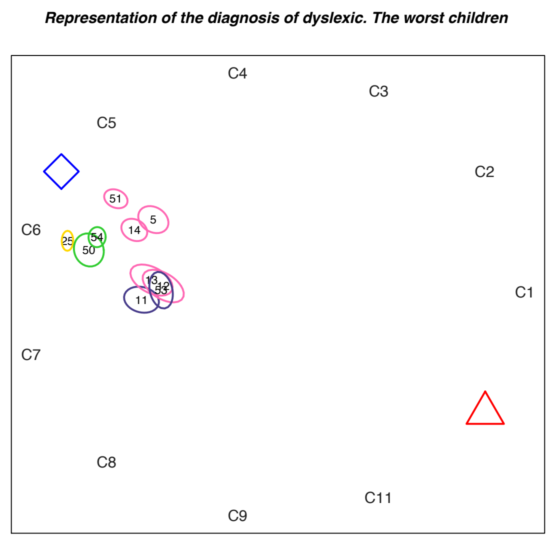

Fig. 16. Worst 10 children with respect to all tests

that allow the supervisor to judge the impact of different techniques into the different categories of students.

From a methodological point of view, an extension of the Radial Coordinates Visualization algorithm (RadViz) to fuzzy data has been used. This technique allows representing individuals whose features are vague and in some cases missing, by means of ellipses fitted to the chosen cuts of the projections of the vague data.

Acknowledgements

This work was supported by the Spanish Ministerio de Economía y Competitividad under Project TIN2011-24302, including funding from the European Regional Development Fund.

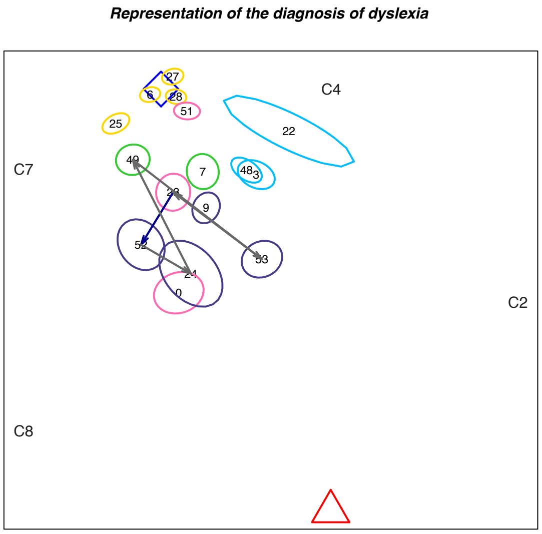

Fig. 17. Position a new child in the classroom.

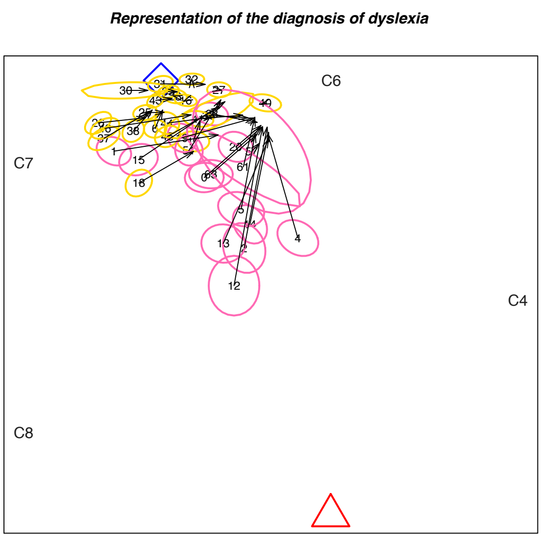

Fig. 18. Neutral evolution of children after a specific learning test.

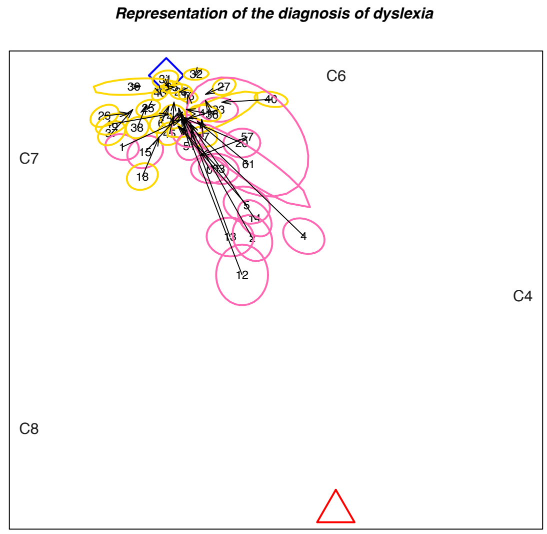

Fig. 19. Positive evolution of children after a specific learning test.

REFERENCES

[1] M. Ankerst, D. Keim,and H.P. Kreigel, *Circle segments: a technique for visually exploring large multidimensional data sets*, in Proceedings of the IEEE Visualization Conference.

[2] U. Fayyad and G.G. Grinstein, *Introduction to Data Visualization*, in: U. Fayyad, G.G. Grinstein, and A. Wierse (eds.), Information Visualisation in Data Mining and Knowledge Discovery, Morgan Kaufmann Publishers, 1-17, 2002.

[3] G.G. Grinstein, M. Trutsch, and U. Cvek, *High-dimensional visualizations*, In Proceedings of Workshop on Visual Data Mining, ACM Conference on Knowledge Discovery and Data Mining, 1-14, 2001.

[4] G.G. Grinstein and M.O. Ward, *Introduction to Data Visualization*, in: U. Fayyad, G.G Grinstein, and A. Wierse (eds.), Information Visualisation in Data Mining and Knowledge Discovery, Morgan Kaufmann Publishers, 21-45, 2002.

[5] H.M Hearnshaw and D.J. Unwin, editors, *Visualization in geographical information systems*, New Brunswick, MJ:Wiley, 1994.

[6] P. Hoffman, G. Grinstein, and D. Pinkney, *Dimensional Anchors: A graphic primitive for multidimensional multivariate information visualizations*, in Proceedings of the Workshop on New Paradigms in Information Visualization and Manipulation, 8^{th} ACM International Conference on Information and Knowledge Management, 9-16, 1999.

[7] R. Kruse, C. Borgelt, F. Klawonn, C. Moewes, M. Steinbrecher, and P. Held, *The Extension Principle,* in Texts in Computer Science, Springer London, 2013, pp. 321-327.

[8] A. Palacios, L. Sánchez, and I. Couso, *Diagnosis of dyslexia with low quality data with genetic fuzzy systems*, International Journal on Approximate Reasoning 51, 993-1009, 2010.

[9] R.M. Rivas and P. Fernández, *Dislexia, disortografía y disgrafía*, Pirámide, 1994 (in Spanish).

[10] A. Salas, E. Gómez, H. Alvarado, M.A. Damians, N. Martorell and S. Sancho, *Protocolos de deteccion y actuacion en dislexia*, Nº de registro: PM-386-2010, 2010.

[11] L. Sanchez, Couso I., Otero J., Palacios A., *Assessing the evolution of learning capabilities and disorderes with graphical exploratory analysis of surveys containing missing and conflicting answers*, Neural Network World 20, 825, 2010.

[12] B. Shneiderman, *The eyes have it: A task by data type taxonomy of information visualization*, presented at IEEE Symposium on Visual Languages, 1996.

[13] A. Stewart, *Trends in quantitative methods III: stressing the visual*, Progress in Human Geography 23, 597-606, 1999.

[14] P.N. Tan, M. Steinbach, and V. Kumar, *Introduction to Data Mining*, Addison-Wesley Longman Publishing Co., Inc., Boston, MA, USA, 1st edition.

[15] K. Techapichetvanich and A. Datta, *Visual mining of market basket association rules*, in: Computational Science and Its Applications - ICCSA Proceedings, Part IV. Lecture Notes in Computer Science 3046, 479-488, 2004.

[16] M.E Thomson, *La dislexia. Su naturaleza, evaluacion y tratamiento*, Alianza, 1992 (in Spanish).

[17] A. Wierse, *What can Visualization do for Data Mining*, Information Visualisation in Data Mining and Knowledge Discovery, 123-124, 2002.