Indonesian Journal of Electrical Engineering and Computer Science
Vol. 24, No. 1, October 2021, pp. 269~278
ISSN: 2502-4752, DOI: 10.11591/ijeecs.v24.i1.pp269-278

Density-based classification with the DENCLUE algorithm

Mouhcine El Hassani, Noureddine Falih, Belaid Bouikhalene
LIMATI Laboratory Polydisciplinary Faculty, University of Sultan Moulay Slimane, Beni Mellal, Morocco

Article Info

Article history:

Received May 17, 2021
Revised Jul 17, 2021
Accepted Aug 4, 2021

Keywords:

Attractor
Clustering
Data mining
DBSCAN
DENCLUE

ABSTRACT

Classification of information is a vague and difficult to explore area of research, hence the emergence of grouping techniques, often referred to Clustering. It is necessary to differentiate between an unsupervised and a supervised classification. Clustering methods are numerous. Data partitioning and hierarchization push to use them in parametric form or not. Also, their use is influenced by algorithms of a probabilistic nature during the partitioning of data. The choice of a method depends on the result of the Clustering that we want to have. This work focuses on classification using the density-based spatial clustering of applications with noise (DBSCAN) and DENsity-based CLUstEring (DENCLUE) algorithm through an application made in csharp. Through the use of three databases which are the iris database, breast cancer wisconsin (diagnostic) data set and bank marketing data set, we show experimentally that the choice of the initial data parameters is important to accelerate the processing and can minimize the number of iterations to reduce the execution time of the application.

Corresponding Author:

Mouhcine El hassani
LIMATI Laboratory, Polydisciplinary Faculty
University of Sultan Moulay Slimane
Mghila, BP 592 Beni Mellal, Morocco
Email: elhassanimouhcine@hotmail.com

1. INTRODUCTION

The desire to classify in order to reduce and better control has gradually developed towards the ambition of automate classification to design and, why not, predict the future. This human vision has pushed scientific community in the field, for years, to improve and find new solutions allowing to release useful knowledge through powerful and autonomous applications which concern automatic classification, simulation and visualization in two or three dimensions of events that may occur in the future. In the analysis of statistical data, an individual is associated with a class among several predefined classes. But in the case of an unsupervised classification, the classes are not known in advance, we then group together into individuals or objects having common properties from a large number of data, hence the complexity of the grouping and identification of the number of classes. Such a classification arose in the analyzes of archaeological data (classifying objects according to age), and medical data (classifying patients according to age, weight, and symptoms).

Subsequently, other uses of classificationin the processing of text data, of image recognition of knowledge extraction [1], which pushed researchers to focus on advanced research of algorithms and clustering techniques as they progress. and with the development of computer tools, hence the emergence of a science aimed at the processes of extracting knowledge from data [2]. In some situations the look and format of clusters can be very useful for analysis and knowledge extraction when visualizing big data and choosing clusters. One can find clusters embedded in others or isolated, depending on the representation of the data. In this context, the contribution we make is the realization of a data mining computer application whichanalyzes data from the preprocessing phase to classification using density-based clustering.

Most recently, Platoš [3] presented a density-based clustering analysis. The idea was to identify dense, fine-grained regions in the data, their grouping produces clusters of arbitrary shape. It is a hierarchical classification algorithm called density-based spatial clustering of applications with noise (DBSCAN) based on density. Groups are formed by dense grid regions of adjacent connectivity, since they share a common side and corner. Connected regions can be found by crossing first or deep using a graph-based model. The algorithm gives good representations with the noise points. The contour of the clusters is smoother, while the rectangular regions are substituted by a spherical area identified by the radius

We approach in our research method two algorithms, the DBSCAN which allows clustering [4] by density, and the DENsity-based CLUstEring (DENCLUE) [5], it was proposed [6] between 1998 and 2000, based on mathematical functions. Although it acquires high complexity with the number of input parameters, it shows acceptable results. Given the importance of the uses of classification in botany, medicine and banking marketing, we tested in section 3 three corresponding databases.

2. RESEARCH METHOD: density-based clustering

Density-based clustering [7], [8] uses the notion of neighborhood [9], [10] to determine a kernel cluster $\mathrm{x_i}$. $\mathrm{N_\varepsilon(x_i)}$ is the neighborhood of $\mathrm{x_i}$ it is the set of points of X whose distance from $\mathrm{x_i}$ is less than or equal to the radius ε:

$$\mathrm{N_\varepsilon(x_i)} = \{\mathrm{x_j \epsilon X |\ d(x_i, x_j) \leq \varepsilon}\} \tag{1}$$

For this type of algorithm, we need to define two parameters: ε the minimum radius around the kernel and M the minimum number of points for the neighborhood $\mathrm{N_\varepsilon(x_i)}$.

The density-based spatial clustering of applications with noise (DBSCAN) [11] algorithm is one of the best known using these two parameters for the identification of clusters based on the notion of neighborhood around a nucleus. The algorithm is very easy to understand and does not require that we provide it with the number of clusters to find. It is able to handle absurd data and get it rid of from the partitioning process. This algorithm can be described as follows Algorithm 1:

Algorithm 1 : density-based spatial clustering of applications with noise (DBSCAN)

```
Algorithm  DBSCAN(Density-Based Spatial Clustering of Applications with Noise)
Input :  two parameters: $\boldsymbol{\varepsilon}$ the minimum radius around the nucleus and $\boldsymbol{M}$ the minimum number
of points for the neighborhood fixed in advance and the set of points $\boldsymbol{X}$ .
Output : Partition $\boldsymbol{C} = \{\boldsymbol{C_1}, \dots, \boldsymbol{C_k}\}$   of $\boldsymbol{X}$ with  k d-clusters.
Begin
      -   Initialize  cluster $\boldsymbol{C_{id}} = \boldsymbol{\phi}$  with  id=1
      -   For i =1   to  n do
              a.  If  $\boldsymbol{x_i}$ is not a core or if $\boldsymbol{x_i \epsilon} \cup_{j=1,\dots,id} \boldsymbol{C_j}$ then go back to step 2
              b.  Build Cluster $(\boldsymbol{x_i}, \boldsymbol{X}, \boldsymbol{C_{id}}, \boldsymbol{\varepsilon}, \boldsymbol{M})$
              c.  $\boldsymbol{id} = \boldsymbol{id} + \mathbf{1}$   and     $\boldsymbol{C_{id}} = \boldsymbol{\phi}$
          End for
      -   Return all the d-clusters: $\boldsymbol{C_1}, \dots, \boldsymbol{C_{id-1}}$
End.
```

Although this algorithm is easy to understand, it remains slow computer execution, especially when the number of points is large. Its complexity is quadratic, of the order of (n^2) [12], [13], but can be reduced to O (nlog (n)) by simplifying the implementation of the algorithm.

Another algorithm also allows density clustering, it is the DENsity-based CLUstEring (DENCLUE) algorithm [14]. Although it acquires a high complexity with the number of input parameters, it has the following advantages:

- Very efficient when dealing with aberrant data presenting noise;
- Capable of mathematically describing arbitrarily chosen clusters belonging to large data sets;
- Fast compared to DBSCAN [15] and therefore stronger.

This algorithm is based on estimation of the density of the kernel through different functions, it is almost the same principle as the DBSCAN algorithm, except that here we have the possibility to illustrate internal hierarchical structure in the distribution data by adjusting window width σ of the kernel influence function.

Let $\mathrm{D} = \{\mathrm{x_1}, \dots, \mathrm{x_n}\}$ denote the data set of n objects in the space Ω. We describe the DENCLUE algorithm by the following notions:

- The estimation of the kernel by the overall density function

  $\forall\ x \in \Omega$, the probability density function [16] is given by:

$$f^D(x) = \frac{1}{n}\sum_{i=1}^{n} K(\frac{x-x_i}{\sigma}) \qquad (2)$$

With K(x) the kernel influence function which is a symmetrical density function with a peak at the origin, it can be a Gaussian function, a square wave function.
σ: the window width of the kernel function.

- Density attractor and density attraction, let x^*denote a local maximum point of the global density function, for a point $\mathrm{x} \in \Omega$. if there exists a set of points $\mathrm{x}_0, \dots, \mathrm{x}_\mathrm{k}$, such that $\mathrm{x}_0 = \mathrm{x}$ and $\mathrm{x}_\mathrm{k} = \mathrm{x}^*$ and x_i $(0 < \mathrm{i} < \mathrm{k})$ so that it lies in the gradient direction of $\mathrm{x}_{\mathrm{i}-1}$, then x is attracted in density by $\mathrm{x}^*$and $\mathrm{x}^*$is an attractor of density of x.

If the kernel function K(x) is continuous and differentiable at each point, the gradient-based escalation method can be used to find the density of attractors.

- Center-based clustering. With x^* a given density attractor, if there exists a subset $\mathrm{C} \subseteq \mathrm{D}$ such that x is attracted in density by x^* and $\mathrm{f}^\mathrm{D}(\mathrm{x}^*\ ) \geq \xi$ with ξ is a noise threshold preset, then C is the cluster with x^* its center.
- Clustering with an arbitrary shape: let X be a set made up of attractor density. If there is a $\mathrm{C} \subseteq \mathrm{D}$ subset which verifies:

$\forall\ \mathrm{x}\ \in \mathrm{C}$, there exists an attractor of density$\mathrm{x}^*\ \in \mathrm{X}$ so that x is attracted in density by x^* and $\mathrm{f}^\mathrm{D}(\mathrm{x}^*\ ) \geq \xi$;
$\forall \mathrm{x}_\mathrm{i}^*, \mathrm{x}_\mathrm{j}^* \in \mathrm{X}$ (i ≠ j), there exists a path $\mathrm{P} \subset \Omega$ from x_i^* to x_j^* which satisfies the following condition:
$\mathrm{y} \in \mathrm{X}, \mathrm{f}^\mathrm{D}(\mathrm{y}\ ) \geq \xi$. C is called the cluster of the arbitrary shape determined by X.

Necessarily, two parameters must be provided to execute this algorithm, which are σ : the window width of the kernel function and ξ the preset noise threshold, the choice of these two parameters influences attractors and number of found clusters.

The basic steps of the DENCLUE Algorithms are as follows:

- Determine the density attractors.
- Associate data objects with density attractors using escalation.
- If possible, merge the initial clusters by relying more on a hierarchical clustering approach.

However, although it takes into account incomplete data and outliers [17], its complexity can be on the order of $O(nlog(n))$, which is acceptable.

We have implemented in our application the steps defined in the DENCLUE algorithm below (Algorithm 2), it receives as parameters the database D and the noise threshold ξ:

Algorithm 2: DENCLUE [18]

```
Algorithm DENCLUE(Dataset: D, Threshold: )
Begin
Determine the density attractor of each data point in a dataset with gradient ascent rule;
Create clusters of data points that converge to the same density attractors;
 Discard clusters whose density attractors have density less than (noise and outliers);
Merge clusters whose density attractors are connected with a path of density at least
return clusters;
End.
```

However, before using raw data, it went through a preprocessing step to eliminate those that contain type errors that could present outliers and delay the classification step and subsequently cloud the visualization of the clusters.

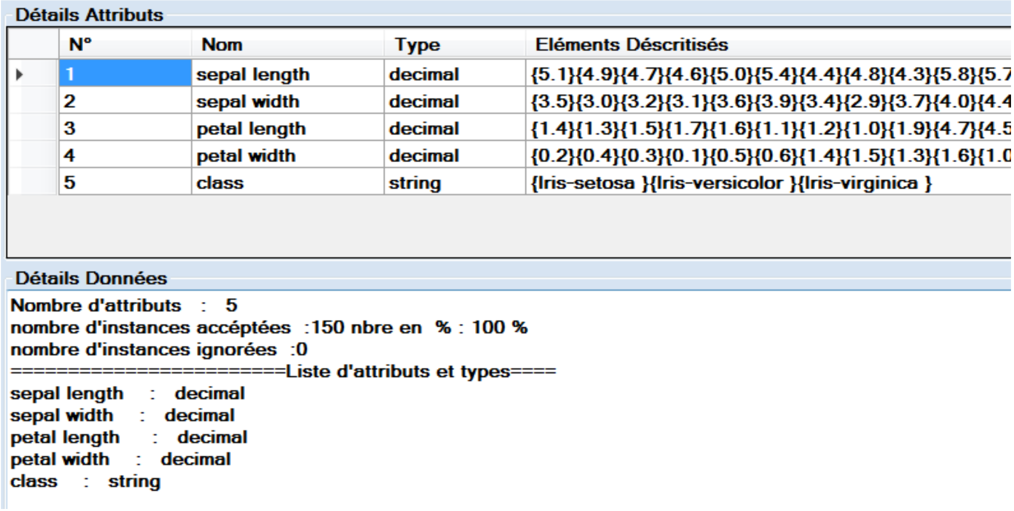
3. RESULTS AND DISCUSSION

The evaluation of the results of a clustering method remains an open problem. But here the main difficulty lies in the fact that the evaluation of the results of the classification is subjective in nature. Accordingly, there are several relevant ways for classifying data objects. In practice, and to verify the reliability of this density-based classification technique, an application has therefore been made in C-Sharp. It groups together several functionalities starting with the preprocessing, then the processing and classification of data.

The application allows user to decide on the choice of initial data and noise threshold. In order to test and evaluate, we use a databases from the web site of an American open access server named “the UC irvine machine learning repository" (UCI) [19]. This server hosts a machine learning repository which is a collection of databases, domain theories and data generators used by the machine learning community for empirical analysis of algorithms.

3.1. Tests on the iris database

Initially we tested the application on a database of which we know the clusters to find. In this context, we use the iris database; it contains 3 classes of 150 records under the following five attributes: length and width of sepals and petals as well as the species. The choice of this base is the best known example in the field of machine learning. The system classifies iris flowers into three species (setosa, versicolor, and virginica) based on measurements of the length and width of sepals and petals. First we apply a preprocessing and processing phase to the data in order to eliminate those containing errors, then we classify them using DBSCAN and DENCLUE. The Figure 1 represents all the data after cleaning. The processing and classification time is very short considering the number and size of this database, it is on the order of a 150 milliseconds. To have a clear view of the nature of the data, the Figure 2 determines the repetitions number for each attribute:



Détails Attributs

N°	Nom	Type	Eléments Déscritisés
1	sepal length	decimal	{5.1}{4.9}{4.7}{4.6}{5.0}{5.4}{4.4}{4.8}{4.3}{5.8}{5.7
2	sepal width	decimal	{3.5}{3.0}{3.2}{3.1}{3.6}{3.9}{3.4}{2.9}{3.7}{4.0}{4.4
3	petal length	decimal	{1.4}{1.3}{1.5}{1.7}{1.6}{1.1}{1.2}{1.0}{1.9}{4.7}{4.5
4	petal width	decimal	{0.2}{0.4}{0.3}{0.1}{0.5}{0.6}{1.4}{1.5}{1.3}{1.6}{1.0
5	class	string	{Iris-setosa }{Iris-versicolor }{Iris-virginica }

Détails Données

```
Nombre d'attributs   :   5
nombre d'instances accéptées  :150 nbre en  % : 100 %
nombre d'instances ignorées  :0
=======================Liste d'attributs et types====
sepal length    :   decimal
sepal width    :   decimal
petal length    :   decimal
petal width    :   decimal
class    :   string
```

Figure 1. Data visualization after cleaning

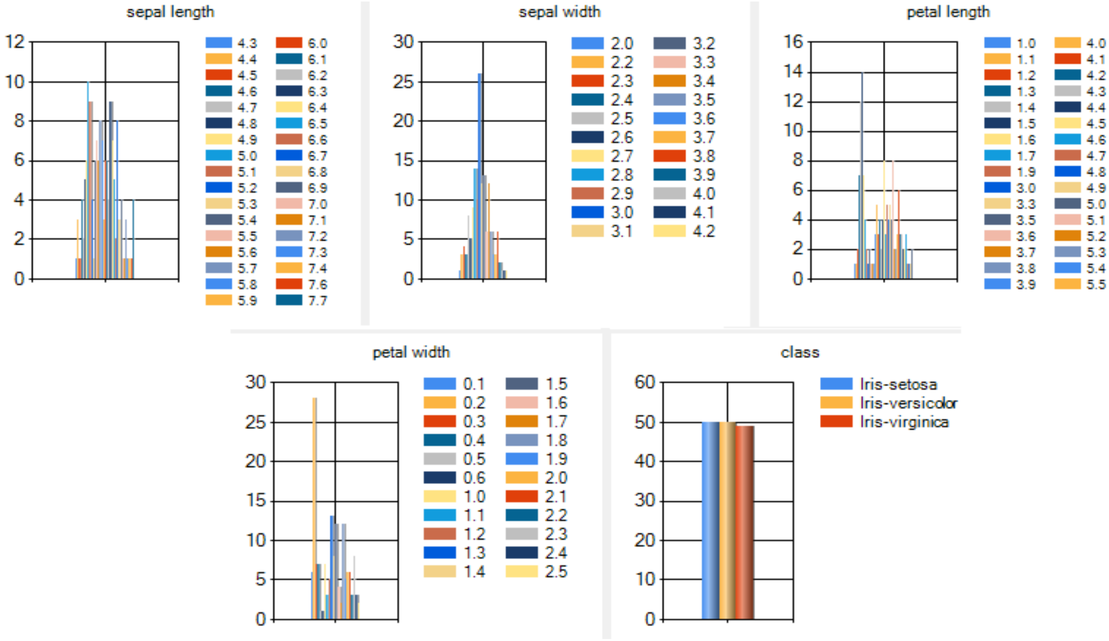


Figure 2. Visualization of the occurrence of data after processing

Finally, for the two algorithms, the choice of initial attributes for the classification is decisive for the identification of clusters. In the Figure 3, clusters can be visible by color of the classes, the crossing of petals length with sepals width gives good result. The three clusters are visible by groupings points of different colors.

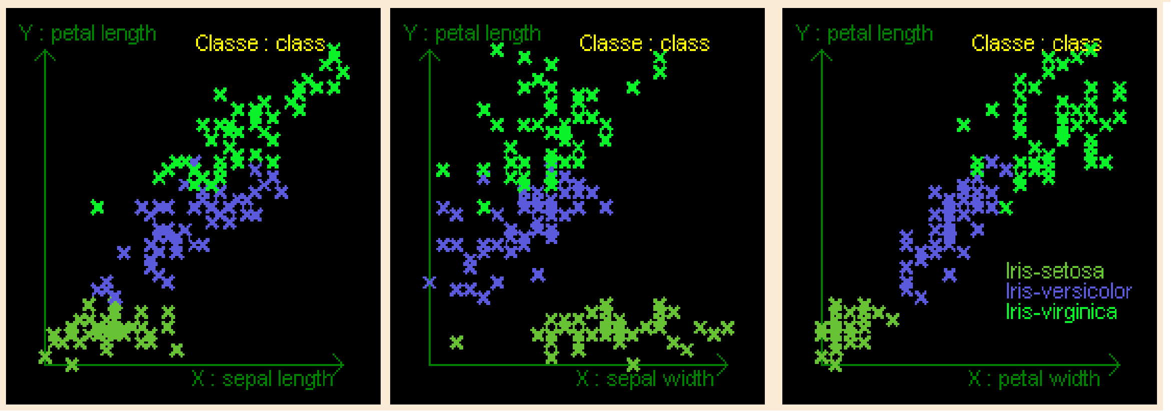


Figure 3. Visualization of iris clusters

3.2. Tests on breast cancer wisconsin (diagnostic) data set

Another test is being applied to the wisconsin hospital breast cancer disease database. As shown in Table 1, this database contains 683 rows and 11 columns. The interest of this choice is to identify the parameters determining the type of breast cancer in order to better target the dosage adopted for patients. After cleaning the data, we presented their occurrences for the set of attributes in Figure 4.

Table 1. Database structure

Attribute	Domain
Sample code number	id number
Clump Thickness	1-10
Uniformity of Cell Size	1-10
Uniformity of Cell Shape	1-10
Marginal Adhesion	1-10
Single Epithelial Cell Size	1-10
Bare Nuclei	1-10
Bland Chromatin	1-10
Normal Nucleoli	1-10
Mitoses	1-10
Class	(2 for benign, 4 for malignant)

The execution of the two classification processes is quite fast, clusters are visible for certain graphs, which makes it possible to identify the relevant attributes, the shapes of observed clusters are not regular, there are also some outliers, but objects with malignant cancer are most common and more dominant. Just below a global visualization of all patient parameters is given by our classification application. An enlarged representation of two graphs is observed, patients suffering from the malignant type concert are visualized by red dots and those of the benign type by white dots as shown in Figure 5.

According to the Figure 4-5, the shape of the clusters can be useful for the validation of the initial parameters to be used in the classification process, we find as significant attributes: clump thickness, uniformity of cell shape and size, mitoses. However, these elements cannot be analyzed with human eye, given the large size and complexity of the information contained therein, hence the need to enrich the database to better visualize isolated cases.

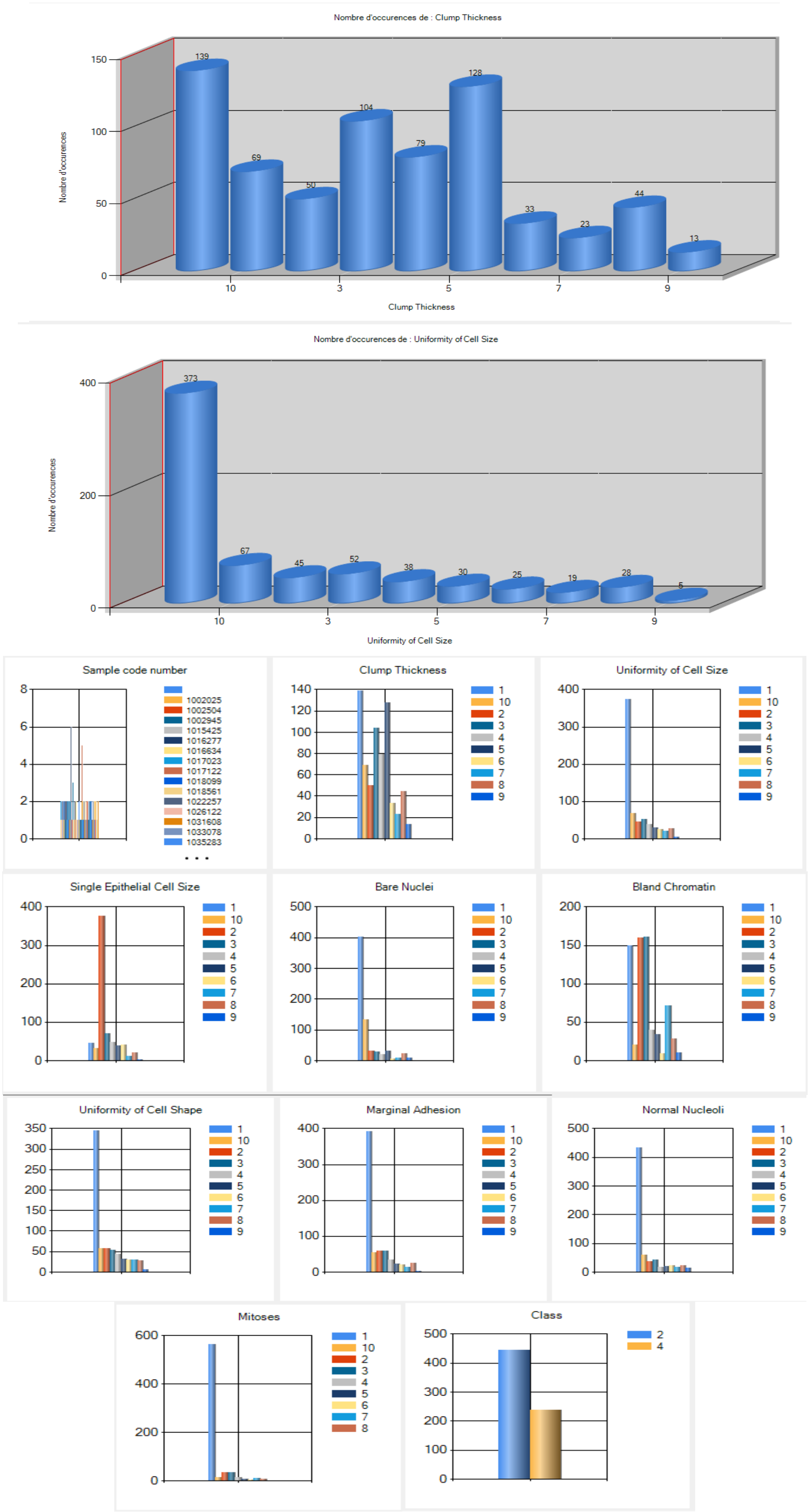
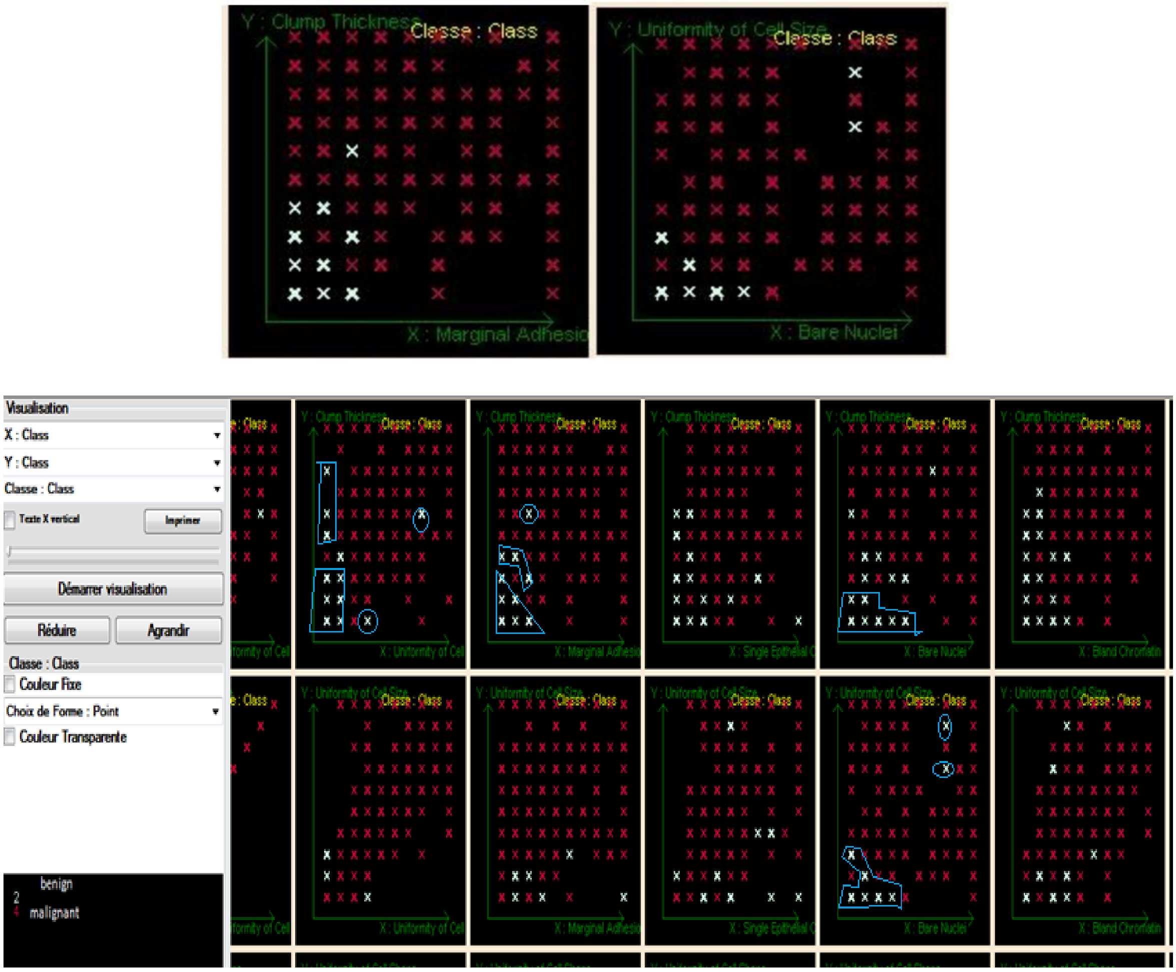


Figure 4. Visualization of the occurrence of data after processing



Visualisation

X : Class

Y : Class

Classe : Class

☐ Texte X vertical

Imprimer

Démarrer visualisation

Réduire

Agrandir

Classe : Class

☐ Couleur Fixe

Choix de Forme : Point

☐ Couleur Transparente

benign

malignant

Figure 5. Visualization of cancer clusters (benign, malignant)

3.3. Tests on bank marketing data set

This database contains about 45,000 rows and 17 columns, so it is important to choose data and number of attributes to enter into the system. As attributes of this database we cite: age, job, marital status and desired target (has the client subscribed a term deposit? "yes", "no"). Given the size and the number of data in this database, we have only presented the following three graphs as shown in Figure 6.

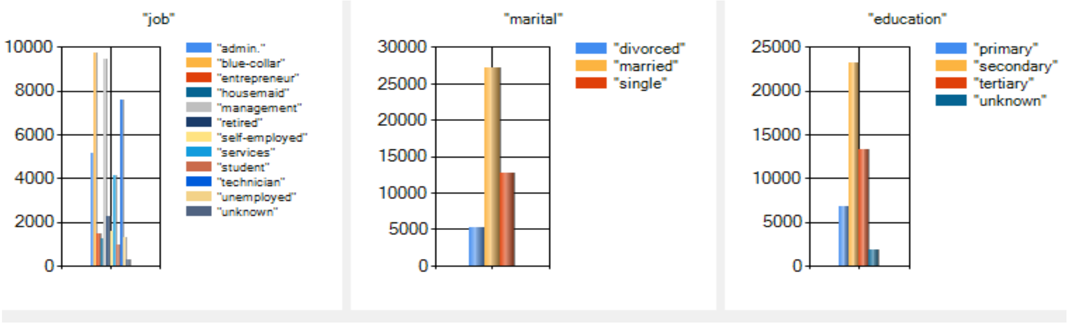


Figure 6. Visualization of occurrences (type of work, martial state, level of study) of individuals

We can see some clusters in different forms and in some cases, they are too close to each other, hence the difficulty of surrounding them as shown in Figure 7. After fixing initial parameters to better calculate density attractors, we construct points with same density which converge. Running system produces a grouping of points in non-regular shapes. Clusters can be seen with the naked eye, but there are also outliers that cloud their visibility. In addition, the choice of the noise threshold and the size of data slow down the processing and produces exceptions related to the limits performance of the computer hardware

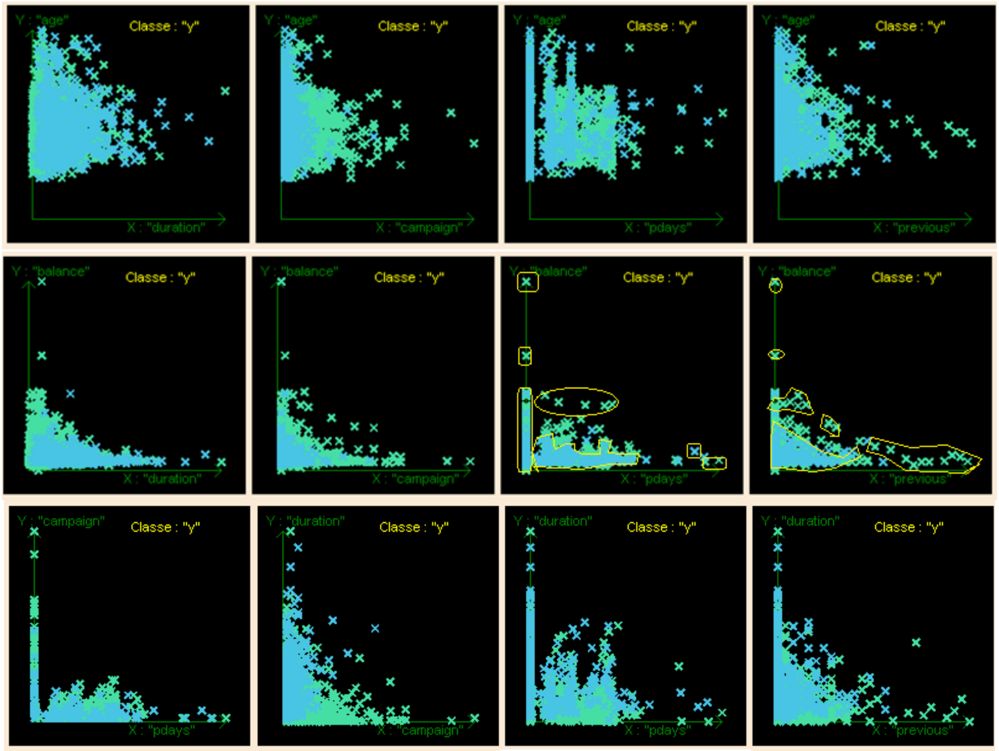


Figure 7. Visualization of bank marketing data set

3.4. Discussion and evaluation

The evaluation of the method [20] can be done on the basis of the found clusters and the quality of knowledge [21]-[23] extracted for good decision-making. However the choice of the density of the attractors and the noise threshold [24] are determining factors for a good measurement of evaluation. Escalation during the search for density attractors [25] and the size of the data to be processed intervene in evaluation of the method. In the example of database (bank marketing data set), the waiting time has been increased from 10 minutes until the system crashes, depending on the number of attributes and rows initially chosen.

The Table 2 shows that the DENCLUE algorithm is better for large databases, it allows choosing input parameters to reduce execution time. In addition, when making a choice of initial parameters to speed up the process, it is possible to omit identification of some of the outliers of interest in decision making. To assess the quality of clustering, we consider how compact the clusters are and how separated they are, it depends on what we want as a result of clustering. The measurement of the average width of the silhouette [26] of the clusters allowed us to present the quality of the clusters in the table.

Another way of evaluation is to call in an expert [27] to understand the meaning of grouping in a particular field. However, while it is possible for an expert to tell whether a given clustering grouping makes sense, it is much more difficult to quantify its interest, or to say whether a given result [28], [29] is better than another. Furthermore, the applicability of the method cannot be extended to other types of data.

Table 2. Comparative analysis of the two density based algorithms

Algorithm	Complexity	Cluster format	Parameters	Noise resistant	Cluster quality	RunTime
DBSCAN	O(n2)	Arbitrary	No input parameter	Well	10%	200 ms, infinitely diverges for large data
DENC-LUE	O(nlog(n))	Arbitrary	Depending on the size of database	Better	20%	100 ms better than DBSCAN and increases for data> 10,000

4. CONCLUSION

We have highlighted the classification using the notion of neighborhood for computation the density of neighboring points. We have done tests on three databases to verify the validity of algorithm, the results can be accepted and used under certain conditions. This was motivated by the reduction, as much as possible, of execution time of calculations and better exploitation of data. We had, therefore, succeeded in the realization of an application regrouping the various processes of extraction knowledge from data, passing by the preprocessing to the search of incoherent and incomplete data, towards treatment and decision-making on the choice of initial values for process learning, after saving the prepared data without affecting the initial ones. Among the short-term perspectives, it is advisable to plan to develop the DENCLUE algorithm in a new version thus allowing a better classification of big data and a reduced execution time, without forgetting the outliers which can be decisive in certain situations.

5. FUTURE WORK

For future work, we want to develop and optimize our system to allow users to move around a data space and get to where they need to be while being subject to the constraints imposed by the decision-making system. Thus we will try to reduce the execution time of the two algorithms DENCLUE and DBSCAN and to improve the quality of the images representing satisfactory classification results and comparable to those of WEKA, the learning and data mining software of the University of Waikato in New Zealand. Also we will try to have a general or detailed vision with color and shape options that can simplify the visualization of the crossed data and even a storage in high resolution photo files, which constitutes an additional point for the knowledge extraction.

ACKNOWLEDGEMENTS

We would like to thank members of the Laboratory of Innovation in Mathematics and Applications and Information Technologies (LIMATI) laboratory of the poly disciplinary faculty of Sultan Moulay Slimane University in Béni Mellal as well as all those who give us their comments and suggestions in order to improve this work in the future, without forgetting those who will contribute to its publication.

REFERENCES

[1] K. Adnan, R. Akbar, “An analytical study of information extraction from unstructured and multidimensional big data,” *Journal Big Data, Publisher: Springer Nature*, vol. 6, no. 1, pp. 5-15, Oct 2019, doi: 10.1186/s40537-019-0254-8.

[2] Jean-Yves Prax, “Le guide du Knowledge Management. Concepts et pratiques du management de la connaissance,” *Paris, Dunod*, 2002.

[3] Jan Platoš, “Data Analysis 2, Density-based Clustering,” *Department of Computer Science Faculty of Electrical Engineering and Computer Science VŠB*, Technical University of Ostrava, pp. 14-16, 2017.

[4] S. Barman, H. Gope, M. M. Islam, M. Hasan, and U. Salma, "Clustering Techniques for Software Engineering," *Indonesian Journal of Electrical Engineering and Computer Science (IJEECS)*, vol. 4, no. 2, pp. 465-472, November 2016, doi: 10.11591/ijeecs.v4.i2.pp465-472.

[5] A. Hinneburg and H. Henning Gabriel, “DENCLUE 2.0: Fast Clustering based on Kernel Density Estimation,” *Institute of Computer Science Martin-Luther-University Halle-Wittenberg*, Germany, pp 3-11, 2004, doi: 10.1007/978-3-540-74825-0_7.

[6] H. Alexander and D. A. Keim, "An efficient approach to clustering in large multimedia databases with noise," *Proceedings of the Fourth International Conference on Knowledge Discovery and Data MiningAugust*, 19998, pp. 58-65.

[7] J. Lu and Q. Zhu, "An effective algorithm based on density clustering framework," *IEEE Access*, vol. 5, pp. 4991-5000, 2017, doi: 10.1109/ACCESS.2017.2688477.

[8] C. Xiaoming, L. Wanquan, H. Qiu, and L. Jianhuang, “APSCAN: A parameter free algorithm for clustering,” *Pattern Recognition*, *Letters*, vol. 32, no. 7, pp. 973-986, 2011, doi: 10.1016/j.patrec.2011.02.001.

[9] Sébastien Ferré, “Concepts de plus proches voisins dans des graphes de connaissances,’’ *IRISA/Université de Rennes 1 Campus de Beaulieu*, pp. 2-10, 2017.

[10] D. Yu, G. Liu, M. Guo, X. Liu, and S. Yao, "Density peaks clustering based on weighted local density sequence and nearest neighbor assignment," *IEEE Access*, vol. 7, pp. 34301-34317, 2019, doi: 10.1109/ACCESS.2019.2904254.

[11] K. Mahesh Kumar and A. Rama Mohan Reddy, "A fast DBSCAN clustering algorithm by accelerating neighbor searching using Groups method," *ELSEVIER Journal Pattern Recognition,* vol. 58, pp. 39-48, 2016, doi: 10.1016/j.patcog.2016.03.008.

[12] R. Xu and D. Wunsch, "Clustering analysis," in Clustering, *Ed. New Jersey: WileyIEEE Press*, pp.1-3, 2008.

[13] B. Pooja, and A. Priyanka, "Comparative Study of Density based Clustering Algorithms," *International Journal of Computer Applications*, vol. 27, no. 11, pp. 421-435, August 2011, doi: 10.5120/3341-4600.

[14] J. Sander, “Density-based clustering,” *In Encyclopedia of Machine Learning,* Springer, pp. 270-273, 2011, doi: 10.1002/widm.1343.

[15] J. Hou and M. Pelillo, "A new density kernel in density peak based clustering," *23rd International Conferenceon attern Recognition (ICPR), Cancu*n, pp. 468-473, Dec. 2016, doi: 10.1109/ICPR.2016.7899678.

[16] D. Li and Y. Du, “Artificial intelligence, with uncertainty,” *Tsinghua University Beijing*, China, Chapman & Hall/CRC Taylor & Francis Group 6000 Broken Sound Parkway NW, Suite 300, Boca Raton, FL 33487-2742, *International Standard Book Numbe*r-13: 978-1-58488-998-4 (Hardcover), pp. 216-224, 2007, doi: 10.1201/9781584889991.

[17] T. Quy, T. Guillaume, and R. Cyril, "Towards Vandalism Detection in OpenStreetMap Through a Data Driven Approach,"*Conference Paper*, August 2018, pp. 6-7, doi: 10.4230/LIPIcs.GISCIENCE.2018.61,

[18] J. Platoš, "Data Analysis: Density Based Clustering, Cluster Validation," *Department of Computer Science, Faculty of Electrical Engineering and Computer Science*, VŠB, Technical University of Ostrava, October 6, 2020.

[19] D. Dua and C. Graff, “{UCI} Machine Learning Repository,” *University of California, Irvine, School of Information and Computer Sciences*, 2019.

[20] G. H. Shah, C. K. Bhensdadia, and A. P. Ganatra, "An Empirical Evaluation of Density-Based Clustering Techniques," *International Journal of Soft Computing and Engineering (IJSCE),* ISSN: 2231-2307, vol. 2, no. 1, pp. 216-223, March 2012.

[21] K. W. Al-Ani, F. B. Abdullah, and S. Yossuf, "Unequal clustering in wireless sensor network: a review," *Indonesian Journal of Electrical Engineering and Computer Science (IJEECS)*, vol. 22, no. 1, pp. 419-426, April 2021, doi: 10.11591/ijeecs.v22.i1.pp419-426.

[22] P. H. Ahmad and Shilpa Dang, "Performance evaluation of clustering algorithm using different datasets," *International Journal of Advance Research in Computer Science and Management Studies*, vol. 3, no. 1, pp. 167-173, January 2015.

[23] S. Patel and A. Patel, "Performance Analysis and Evaluation of Clustering Algorithms," *International Journal of Innovative Technology and Exploring Engineering (IJITEE)*, vol. 8I, no. 6S2, pp. 179-183, April 2019.

[24] H. P. Kriegel, P. Kröger, J. Sander, and A. Zimek, "Density-based Clustering," *WIREs Data Mining and Knowledge Discovery*, vol. 1, no. 3, pp. 231-240, 2011, doi: 10.1002/widm.30.

[25] O. Niphaphorn, and S. Wiwat, "Optimal Choice of Parameters for DENCLUE-based and Ant Colony Clustering," *International Conference on Modeling, Simulation and Control, IPCSIT, IACSIT Press*, Singapore, vol. 10, 2011., pp. 71-72.

[26] F. Batool and C. Hennig, "Clustering with the Average Silhouette Width," *Journal of Computational Statistics & Data Analysis*, vol. 158, February 2021, doi: 10.1016/j.csda.2021.107190.

[27] A. Styhre, "Knowledge Sharing in Professions: Roles and Identity in Expert Communities," *Publisher book: Routledge* in 2016, pp. 83-153, ISBN 1409420973, doi: 10.4324/9781315591193.

[28] S. M. Mohammed, K. Jacksi, and S. R. M. Zeebaree, "A state-of-the-art survey on semantic similarity for document clustering using GloVe and density-based algorithms," *Indonesian Journal of Electrical Engineering and Computer Science (IJEECS)*, vol. 22, no. 1, pp. 552-562, April 2021, doi: 10.11591/ijeecs.v22.i1.pp552-562.

[29] N. Sharma, A. Bajpai, and R. Litoriya, "Comparison the various clustering algorithms of weka tools," *International Journal of Emerging Technology and Advanced Engineering,* vol. 2, no. 5, pp. 74-79, May 2012.