Unsupervised Classifier for DynamicSystems Based on ISODATA

Cela A.*; Sánchez L.**

**Escuela Politécnica Nacional, Facultad de Ingeniería Eléctrica y Electrónica*
Quito, Ecuador (Tel: 593-2-2677-031; e-mail: andres_cela@ieee.org)
***Universidad Carlos III de Madrid, Facultad de Ingeniería Telemática*
Instituto Tecnológico Superior Sucre, Departamento de Electrónica
Madrid, España (Tel: 34-6-00312-583; e-mail: franklin.sánchez@alumnos.uc3m.es)

Resumen:*Elsiguiente trabajo describe el método empleado para diseñar un algoritmo que permita clasificar un conjunto dinámico. Aunque hoy en día los clasificadores son muy bien conocidos, este trabajo muestra la combinación de diferentes técnicas para obtener un clasificador de mejores prestaciones aplicado a un objetivo específico. Los elementos de estudio pertenecen a un conjunto de líneas las cuales se representan con dos parámetros y se han obtenido después de tratar laimagen de una carretera y obtener todas sus líneas. El primer parámetro corresponde a la pendiente de la línea y el segundo a la posición (x,y) dentro de la imagen. Las imágenes se han obtenido de un recorrido real por una carretera, razón por la cual los elementos, es decir, las líneas, van cambiando de posición entre cada imagen, por tanto sus parámetros son dinámicos. Se parte del clasificador básico de Bayes para obtener un clasificador tipo KNN. En los elementos de estudio no se puede determinar las clases ni obtener sus parámetros, por lo tanto es necesario aplicar un clasificador no supervisado basado en KNN. En la fase final se hace una modificación en la semilladel clasificador ISODATA para aplicarlo al conjunto dinámico. Para reducir el trabajo de clasificación se trabaja únicamente con dos conjuntos los cuales correspondena las líneas izquierda y derecha de la carretera respectivamente. Se demuestra que el algoritmo propuesto reduce el tiempo de ejecución en 25% e incrementa los aciertos a un 90% respecto del clasificador ISODATA sin semilla inicial.*

Palabras clave:*clasificador no supervisado, conjuntos dinámicos, estudio de clasificadores.*

Abstract:*This paper describes a method for designing an algorithm to classify a dynamic set. Nowadays, the classifiers are well known, this work shows a combination of different techniques in order to get a better classifier for a specific work. The set of elements correspond to lanes marks of a road and they have been obtained from a road image. The lines have two parameters which are the slope line and the (x,y) position of the line in the image.The lines are obtained for each frame of a video sequence. The lines position changes between frames, then the parameters also change, this mean they are dynamics .The Bayes theory is the initial point to get a KNN classifier. In the set of study there is not possible to calculate the classes or their parameters, then, an unsupervised classifier based on KNN classifier is used for this approach. Finally, in the last step, it is made a modification in the seed for applyingin a dynamic set. It is shown that this approach reduces the process time in 25% and improves the accuracy in 90%.*

Keywords:*unsupervised classifier, dynamic sets, classifiers.*

1. INTRODUCTION

The classification of elements from a population has well been studied for the last two decades. The classifier calculates the probability that an element belongs to one category or another.

A category, called also class or pattern, can represent different things, like a car, an animal, a shape, a face, etc.

Categories can have different features, commonly called dimensions.These categories have the characteristics of the pattern, like color, texture, dimensions, position, velocity, etc. It is easy to think that to have more features is better for the classifier, but that is not true. Therefore, there are complete studies to obtain the better features which allow to classifythe population. To select the minimum features is recommended by some researchers in [1] and [2]. The categories belong to the same set and same features but each one is located in a different place in the d-dimensional space, where d is the number of features.

Some systems are static, but there is also some dynamic population. In these systems the categories change in time[3].

Therefore, a different kind of classifier is applied. The proposal in this work is to use an unsupervised classifier in order to do the classification based on ISODATA Clustering (IDC) method[4]. The data set were collected from a process which obtains images of the road, then, different artificial vision algorithms obtain lines from the road. These lines can belong to the marks of the road lanes. The goal of this

approach is to classify these lines in two categories, the left lane mark and the right lane mark in order to get two estimated lanes marks using the image information.
However, this system is not static but dynamic. This means that, the position of the categories is changing in time.
Classification of dynamic systems has not had many applications because there are other algorithms, like trackers based on Kalman filters, particle filters[5], or other ones[6], that can also do similar classification in these systems.
On the other hand, there are many classifiers, parametric or non parametric, unsupervised or supervised for static systems[7], but for dynamic systemstracking is most used.
This paper shows that an IDC algorithm can classify this kind of systems too.
Next section describes the method to design the classifier. It includes an introduction about Bayesian decision theoryand its application in multidimensional systems. The simplification of the Bayes theory based on the Gaussian shapes is important in order to do the formulation simpler.
Section 3 shows the results in two cases, the first one is applying the basic IDC and the second one is applying the improved IDC.Conclusions and future works are presented in section 4.

2. CLASSIFIER DESIGN

2.1 Bayesian decision theory

The Bayesian decision theory is a fundamental approach for classifying population in categories. It is based on quantifying the probability of each element to belong a category[4]. Also, it uses probabilistic terms and for each category the probability can be measured direct or indirectly. There is a basic rule which allows to find this probability. If avariable x, or element that belongs to a set C, satisfies (1), then it will be classified in class w_1, instead class w_2 because the probability that belongs to w_1 is higher [7].

$$p(x\,|\,w_1) > p(x\,|\,w_2) \tag{1}$$

Bayes' formulais the general way to understand this theory. It converts the prior $P(w_i)$ probability to the posterior probability $P(w_i\,|\,x)$. The $p(x\,|\,w_i)$ function is the likelihood of w_i respect to x and i is the number of categories[8], as it is described in (2).

$$P(w_i\,|\,x) = \frac{p(x\,|\,w_i).P(w_i)}{p(x)} \tag{2}$$

If (2) is combined for two categories it will obtain (1). The evidence factor $p(x)$ is the same for all categories and can be viewed as a merely scale factor to normalize to 1 the posterior probability.
In some cases the prior probability can be observed like a modulation factor, but if this is not known it can be calculated by using (3), where $i_{\max}$ is the total number of categories.

$$P(w_i) = 1/i_{\max} \tag{3}$$

The prior probabilities modulate the $p(x\,|\,w_i)$ functions. However, it is important to know the prior probability because it can give more weight or cost to any category.
Generally, the Gaussian shape, called also Normal density, is used to study the basic classifiers. The Probability Density Function (PDF) is represented by $p(x\,|\,w_i)$ and it will be defined by a Gauss function. In the Fig.1 the PDF of each category is multiplied by the prior probability, whichis why the Gaussians have different altitudes.

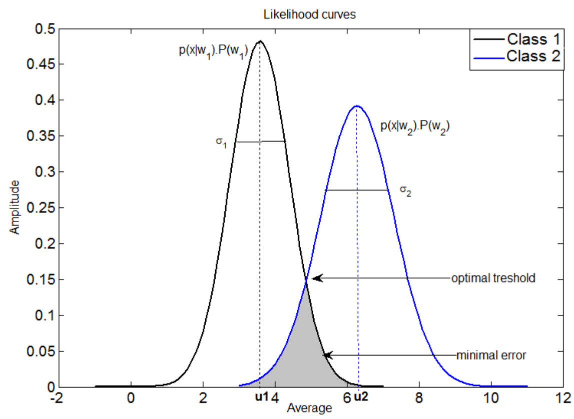


Figure 1.Decision theory.

The normal shape function for univariateis defined by (4) where the probability is calculated by the normal function. In the mono variable function u is the mean and σ is the variance. Therefore, the mean and the variance are the parameters which specify the normal function[9].

$$p(w\,|\,\sigma) = 1/(\sigma(2\pi)^{1/2}).\exp(-(x-u)^2/(2\sigma)) \tag{4}$$

2.2 PDFmultidimensional simplification

The normal function parameters can be found by different parametric techniques like Maximum Likelihood (ML), Bayesian Learning (BL) or by non parametric techniques like Parzen Window (PW) and K-Nearest Neighbors (KNN)[4].
In this work the population is a dynamic system, so the parameters of each category are dynamic too.
Applying a logarithmic function to (4) it will be simplified as it is shown in (5).

$$\ln(\theta(u,\sigma)) = \ln(u,\sigma) = -\ln(2\pi)^{1/2} - (x-u)^2/(2\sigma) \quad (5)$$

ThePDF for multidimensional systems is defined by (6), where $\overline{x}$ is the n-dimensional variable, $\overline{u}$ is the mean vector and Σ is the covariance matrix.

$$L(\overline{u},\Sigma) = -Ln((2\pi)^{1/d}\,|\Sigma|^{1/2}) - (\overline{x}-\overline{u})^T\Sigma^{-1}(\overline{x}-\overline{u})/2 \quad (6)$$

With this transformation it is possible to see that the normal shape implied to use the distance between the variable and the mean, as it is showed by the factor $(\overline{x}-\overline{u})^T\Sigma^{-1}(\overline{x}-\overline{u})$. This is known as the Mahalanobis Distance (MD)[9].

It is assumed that the variables do not have correlation between them and the covariance matrix is an identity matrix. Then, the MD became Euclidean Distance (ED). Besides, the first factor in (6) is a constant for all variables and it can be eliminated. Thereby the final PDF is

$$L(\overline{u}) = (\overline{x}-\overline{u})^T I(\overline{x}-\overline{u}) \quad (7).$$

The normal shape hasyielded a PDFbasedon distances.

However, this technique is already a parametric technique and it is necessary to know the static $\overline{u}$, but it has been defined that this will be a dynamic mean instead.

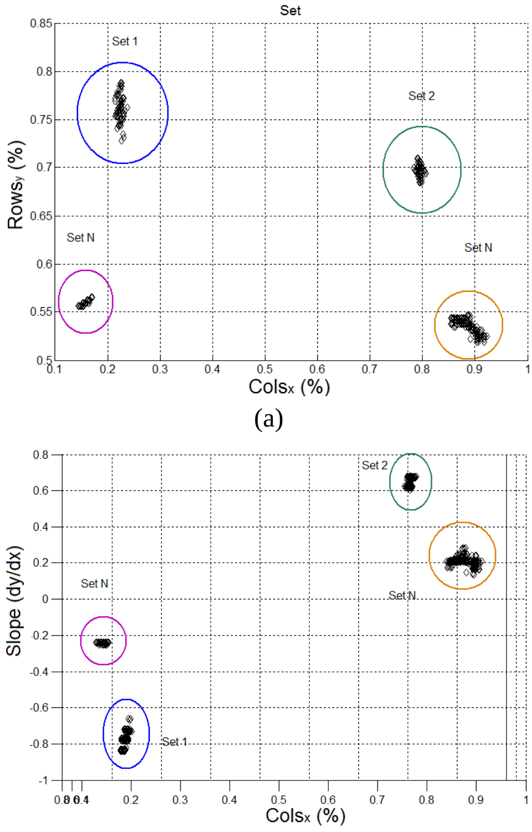
In this paper it is proposed to use an unsupervised classifier to find the dynamic mean.

2.2Finding the mean $\overline{u}$.

The population in this research will start in a known space. There are 4 categories in this set, but only thetwo central categories are considered for classification, as it is shown in Fig.2, where set 1 and set 2 are these categories. These have been considered because their represent the left lane mark and the right lane mark of respectively. The others categories, represented by set N, correspond to lanes marks of others lanes of the road because the road had 3 lanes.

The population is in the 3D space, where X axis represents columns, Y axisrepresents rows and Z axis represents the slope. The population is a set of lines which has been obtained from a sequence of video of road lanes and represent possible lane marks on the road. The (x,y) position is the line location on the image, and the Z is its slope.

The variables, the population, can seem an easy set to classify, but it is necessary to remember that it is only the first frame of a dynamic system and then the categories will change their position on the 3D space. Additionally, the mean of each category is not known yet. This is the aim of this work.



(a)

(b)

Figure 2.Population. (a) (x,y) plane (b). (x,z) plane.

Equation (7) is used for KNN. In this kind of classifier the variable is assigned to the nearest category [10]. However, the means are not known yet and they have to be calculated by another method.

The unsupervised classifier ISODATA Clustering (IDC) is usedto solve this problem. In this method it is not necessary to know the categories because it uses the KNN spectral distance to define the clusters and to find their means [11].

The IDC is an iterative method and it starts with random categories.The KNN classifies the population in the categories defined in the previous step. Then, the means of the categories are calculatedand new means are founded. Whole process is repeated until satisfying a threshold defined by the programmer[11].

Table 1. describes the IDC process.

Table1.ISODATA Clustering process.

1 Cluster centers are randomly placed. $\bar{u}_1$, … ,$\bar{u}_c$.
2 Simple classify by KNN.
3 Calculating new means $\bar{u}$ for each category.
4 If $\Delta\bar{u}$ > min. → go to step 1
Else,Stop.

In this research a data set has been classified and the results are shown in Fig. 3. In some cases the classifier converges to false categories. These results are presented in the next section.

This process calculates the dynamic parameters for categories, in this case the parameter is the mean $\bar{u}$.

3. RESULTS

3.1 ISODATAapplied to a dynamic system.

Fig.2 shows the classification of the first set, but the category means change in time.Hence, the IDC is applied to each frame. The KNN has much delay because it measures the distances for each variable. Besides, if the IDC does not converge in the first frame the whole process will fail.

To solve this issue it is proposed to use a prior mean in step 1 of IDC process. Then, the means obtained in the last step will be the prior mean for the next frame of the dynamic set.

The real lane marks have initial positions in the image which is taken from the first image. To define the first $\bar{u}$ the mean in the expected categories has been measured, which always starts in the same regions with means defined by (8) and (9) which represent the initial positions of categories in the image.These is the initial seed for using the IDCclasiffier.

$$\bar{u}_{1\_0}(x,y,z) = (0.2*x\max, 0.8*y\max, -y\max/x\max) \quad (8)$$

$$\bar{u}_{2\_0}(x,y,z) = (0.8*x\max, 0.8*y\max, y\max/x\max) \quad (9)$$

The result to applying IDC without the prior parameters (8) and (9) to population is depicted in Fig. 3 where the blue elements have been classified in set 1, but some of these elements do not belong to set 1, these are positive false. The red ones have correctly been classified in set 2. The initial seeds u_1 and u_2 have been randomly selected. We have tested this process 200 times and it fails around 60 %.

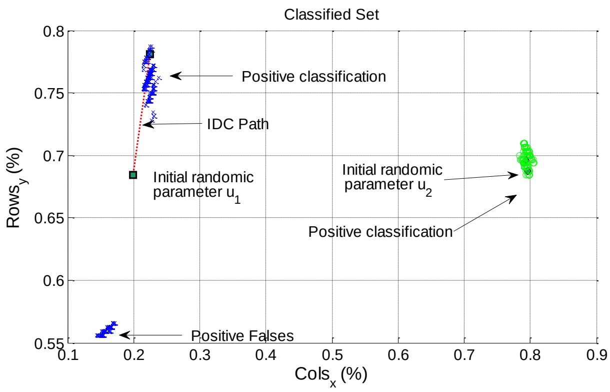


Figure3.ISODATAClustering without $\bar{u}$ prior.

On the other hand, in Fig. 4 is depicted the IDC results with prior parameters. The blue elements have been classified in set 1 and the green ones in set 2. The initial seed in category 1 is $\mathbf{u_{10}}$ which change to $\mathbf{u_{11}}$. The initial seed in category 2 is $\mathbf{u_{20}}$ and the recalculated seed is $\mathbf{u_{21}}$. The black elements have not been classified in any set.In this case IDC has 99% reliability.

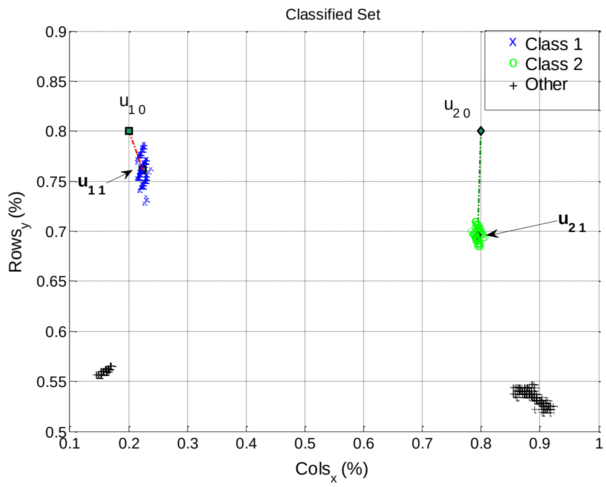


Figure 4.ISODATAClusteringwith $\bar{u}$ prior.

Fig. 5 shows the convergence of four means which started in a random position. Insecond iteration, x1 and x4 means have been changed, and in iteration 3 the means have converged.

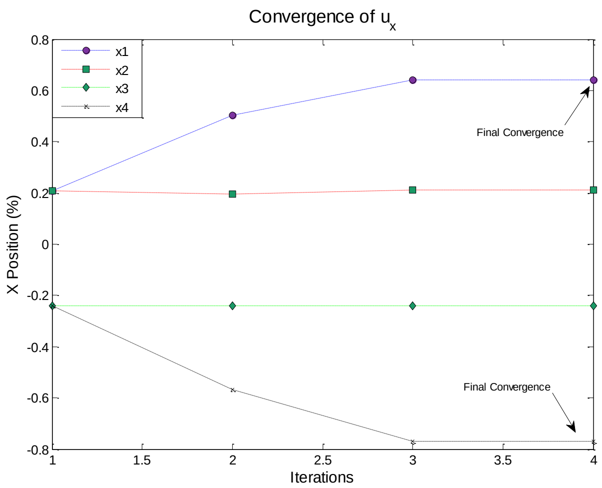


Figure 5.Convergence of means in IDC without initial seed.

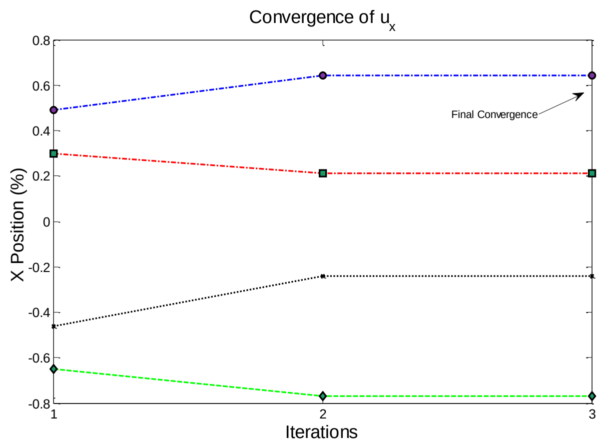


Figure6.Convergence of means in IDC for dynamics sets.

Fig. 6 shows the convergence of the meansusing IDC with initial seed. In iteration 1 the means started in the initial seed. In iteration 3 the means have converged. This method needs just 3 iterations, while typical IDC needs 4.If our method is applied to the whole dynamic system the classification will be 25% faster.

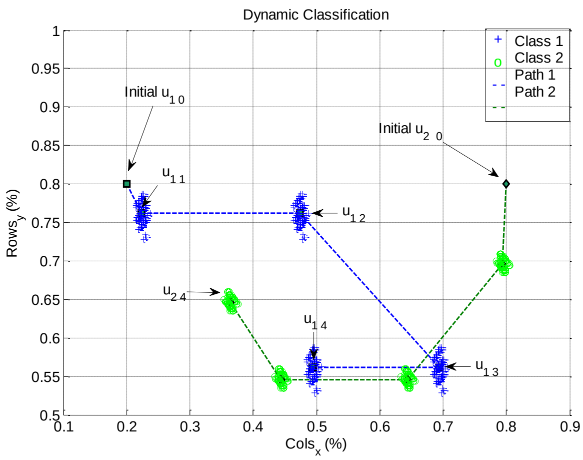


Figure 7.Classifier applied to dynamic system.

The classification of 3D dynamic systems is depicted in Fig.7.Here, the first frame is classified based on the prior parameters (8) and (9). Then, IDC classifies the population in two categories, in green and blue respectively, and it obtains the newer means $\bar{u}_{1\_1}$ and $\bar{u}_{2\_1}$. This algorithm is recurrent until the last frame of population. The blue line and green lines are the paths taken by category 1 and 2. This example shows a classification of a 3D dynamic system for the first 4 frames.

This process can resemble a tracking mechanism, but the difference is that in this case the categories are classified for each frame.

Table 2 shows the algorithm of the improved IDC process applied in dynamic systems. The***k*** variable is the frame number.

Table2.ISODATA Clustering process for dynamic systems.

1 For ***k*** == 0

Cluster centers $\bar{u}_{1\_0}$ = (8) and $\bar{u}_{2\_0}$ = (9).

2 For ***k***> 0 **until** whole frames.

3 Clustercenters placed at: $\bar{u}_{1\_k-1}$ and $\bar{u}_{2\_k-1}$.

4Simple classify by KNN.

5Calculating new means $\bar{u}_{1\_k}$ and $\bar{u}_{1\_k}$.

6If $\Delta\bar{u}_{1\_k}$> min&& $\Delta\bar{u}_{k\_2}$> min. → go to step3

Else →go to step 2, k= k+ 1

4. CONCLUSIONS ANDFUTURE WORKS

The Bayesian classifier is the basis for understanding the scope of decision theory.

An unsupervised classifier allowsto findthe categories in sets.Thiskind of classifier does not require training or supervision, but in some cases the algorithm does not converge to expected categories. In this paper we have used an unsupervised classifier called ISODATA IDC. We have tested IDC with the sets in this paper and it fails around 60%, as we have shown in the results. IDC uses a randomic initial means but we have used an initial seed in order to improved IDC. For each image the improved IDC is 25% faster than the typical IDC and has 99% reliability.

The sets in this work change in time and categories are dynamics which can be classified using this new method, but it is important to know the prior parameter, as in this case.

In future works, we will test other kinds of classifiers and filters. It can also apply a Kalman filter to the means to refine the results and tracking the means.

Although this algorithm is applied to specific sets, we will test the improved IDC in other kinds of sets.

REFERENCES

[1] M. B. A. Haghighat y E. Namjoo, «Evaluating the informativity of features in dimensionality reduction methods», en *2011 5th International Conference on Application of Information and Communication Technologies (AICT)*, 2011, pp. 1-5.

[2] Laurens van der Maaten y ic Postma, «Dimensionality Reduction: A Comparative Review». Tilburg University, 26-oct-2009.

[3] D. Gillies y Y. Zheng, «Dynamic Interactions with the Philosophy of Mathematics», *Theor. Rev. Teoría Hist. Fundam. Cienc.*, vol. 16, n.º 3, pp. 437-459, jun. 2008.

[4] R. O. Duda, P. E. Hart, y D. G. Stork, *Pattern Classification*. John, 2nd ed., Wiley & Sons, 2012, pp. 20-200.

[5] J. Carpenter, P. Clifford, y P. Fearnhead, «An Improved Particle Filter for Non-linear Problems», 2004, pp. 2–7.

[6] M. Isard y A. Blake, «Contour Tracking By Stochastic Propagation of Conditional Density», 1996, pp. 343–356.

[7] K. Fukunaga, *Introduction to Statistical Pattern Recognition*, 2nd ed., Academic Press, 1990, pp. 120-300.

[8] Y.-H. Cai, «The comparative study of different Bayesian classifier models», en *2010 International Conference on Machine Learning and Cybernetics (ICMLC)*, 2010, vol. 1, pp. 309-313.

[9] C. Bishop, *Pattern Recognition and Machine Learning (Information Science and Statistics)*. Springer, 2007, pp.100-200.

[10] J. M. Kleinberg, «Two algorithms for nearest-neighbor search in high dimensions», en *Proceedings of the twenty-ninth annual ACM symposium on Theory of computing*, New York, NY, USA, 1997, pp. 599–608.

[11] J. C. Bezdek, «A Convergence Theorem for the Fuzzy ISODATA Clustering Algorithms», *Ieee Trans. Pattern Anal. Mach. Intell.*, vol. PAMI-2, n.º 1, pp. 1-8, 1980.