Selección de características para atributos continuos en tareas de clasificación de actividad física

Features selection for continuous attributes in classification of physical activity tasks

Seleção de recursos para atributos contínuos em tarefas de classificação de atividade física

Enrique V. Carrera
Universidad de las Fuerzas Armadas ESPE, Ecuador
evcarrera@espe.edu.ec

Jefferson Stalin Rodríguez Páramo
Universidad de las Fuerzas Armadas ESPE, Ecuador
jsrodriguez2@espe.edu.ec

Resumen

Los dispositivos móviles contienen diversos sensores con capacidad para enviar datos que se utilizan en la toma de decisiones, un ejemplo es la clasificación de actividad física basada en el uso de acelerómetros y giroscopios. Las señales de los sensores se procesaron previamente aplicando diferentes técnicas que extrajeron un sinnúmero de atributos, los cuales sirvieron para el desarrollo de tareas de clasificación. La optimización de sistemas de clasificación requirió la disminución del número de características de entrada con la finalidad de sintetizar la dimensión de su conjunto y tiempo de aprendizaje. Este artículo empleó métricas de ganancia de información para atributos continuos, que redujeron la incertidumbre y extrajeron únicamente aquellas características más significativas a través de los datos procesados. El análisis de los resultados que se obtuvieron en la clasificación de actividad física usando redes neuronales, mostraron no solamente la disminución de características, sino también un error por debajo del 5 % y la reducción del tiempo de procesamiento en aproximadamente 55 %.

Palabras clave: aprendizaje de máquina, actividad física, selección de características, atributos continuos, ganancia de información.

Abstract

Mobile devices contain different sensors with the ability to send data that are used in decision-making, an example is the classification of physical activity based on the use of accelerometers and gyroscopes. The signals from the sensors were processed previously applying different techniques which extracted a countless number of attributes, which were used for the development of classification tasks. Optimizing systems of classification required the decrease of the number of input features with the purpose of synthesizing the dimension of its set and learning time. This article used metrics of information gain for continuous attributes, that reduced the uncertainty and extracted only those most significant characteristics through the processed data. The analysis of the results that were obtained in the classification of physical activity using neural networks, showed not only the reduction of features, but also an error below the 5% and the reduction of processing time by approximately 55%.

Key words: machine learning, physical activity, features selection, continuous attributes, information gain.

Resumo

Dispositivos móveis contêm vários sensores capazes de enviar os dados utilizados na tomada de decisões, um exemplo é a classificação de actividade física baseada na utilização de acelerómetros e giroscópios. Os sinais dos sensores são processados através da aplicação de diferentes técnicas extraídos inúmeros atributos, que serviram para o desenvolvimento de tarefas de classificação. A optimização do sistema de classificação necessária a redução do número de características de entrada, a fim de sintetizar a dimensão de tempo em conjunto e aprendizagem. Este artigo usou métricas de ganho informações para atributos contínuos, o que reduziu a incerteza e extraídas apenas as características mais significativas através dos dados processados. A análise dos resultados obtidos na classificação de actividade física utilizando redes neurais, não só mostraram diminuição características, mas também um erro inferior a 5% e o tempo de processamento reduzido em cerca de 55%.

Palavras-chave: aprendizagem de máquina, atividade física, seleção de características, atributos contínuos, ganho de informação.

Fecha recepción: Enero 2016 **Fecha aceptación:** Junio 2016

Introduction

Mobile devices contain different sensors that are currently used in various fields and with countless applications around the world (Das, Green, Perez, and Murphy, 2010). "Smart" mobile devices because of its small size, ability to send and receive data and computing power, allow to store information that can be manipulated (Kwapisz, Weiss, and Moore, 2011).

The information collected by these electronic devices provides a significant contribution to the development and monitoring, aspects relating to health care, rehabilitation, medical diagnosis, people safety, among others (Mitchell, Monaghan, and O'Connor, 2013).

The signals that are emitted by the sensors can not be classified with standards algorithms, so in the first instance it must transform the raw data to information, whose processing is more simple depending on the time or the frequency (Weiss and Hirsh, 1998). In this way, gets out of process and removing a specific number of features based on different metrics.

The large number of existing input data increases the processing time (Han, Kamber, and Pei, 2011), this causes that the optimization of systems of classification demands their reduction. To do this it is necessary to use an algorithm that allows the selection of features, in such a way that will synthesize a whole dimension and learning time (Yang y Wang, 2011).

The classification or automatic selection of features is one of the most common tasks where artificial neural networks have proven themselves, since they perform automatic processing of data and are based on the biological nervous system (Isasi and Galvan, 2004).

It is noteworthy that the Artificial Neural Networks, from its appearance and its rapid development have had a significant use as a technology for data mining, it because the technology has attributes for effective and efficient modeling of complex problems (Lu, Setiono, and Liu, 1996).

This research is based on information processing continuous type using metrics to gain information, which by means of an algorithm will be quantized. Accordingly, the feature selection process for continuous attributes in classification tasks physical activity will be possible; that is, the most important characteristics for the classification process will be identified, by reducing uncertainty and get only those most significant.

Selected characteristics must specify the physical activity of a person (walking up or down stairs, sitting, standing, lying down).

It is worth mentioning that the criterion of maximizing the information gain produces a bias towards the attributes that have lots of different values, which solves this problem by using the ratio of profit as separation criteria (Hong, 1997). This measure takes into account both the information gain as the probabilities of the different values of the attributes; in turn, these probabilities are collected by the separation information, which is not more than the entropy of the data set from the values of the attributes.

The results of the classification of physical activity using neural networks, as described later, it shows that by using: information gain, breakpoints for five groups of selection intervals and error rate in each, he was achieved decrease the set of features (561) in 86% (78), so optimization is perceived in the time data processing.

The structure of this article is as follows: in Section 1 the description of materials and methods adopted is shown, which expose experimental development for capturing data, the mathematical description of the proposed algorithm and the process of training network. In section 2 the experimental results and their analysis are displayed; and finally in section 3 the conclusions.

1. MATERIALS AND METHODS

Dataset and sensors

In the market there is a wide variety of mobile devices which have been developed different operating systems such as Apple's iOS and Google's Android. database "Human Activity Recognition Using Smartphones Data Set" (UCI HAR Dataset) Repository of Machine Learning at the University of California, same as working with a Smartphone (Samsung Galaxy S II) was used in this paper placed at the waist. Through its accelerometer and gyroscope embedded, linear acceleration and angular velocity in three XYZ axes it is obtained. The experiments were videotaped to label the data manually. The dataset is divided randomly into two groups. There, 70% (21 people) of volunteers was selected to generate the training data and 30% (9 people) provided test data.

Selected for this database characteristics derived from the raw signals of the three axes accelerometer and gyroscope, which in the time domain were captured at a constant speed of 50 Hertz (Hz) and were sampled with sliding windows wide fixed 2.56 seconds (s) and 50% overlap (128 readings / window). The acceleration signal has two components: gravitational and body movement; which they are separated and refined in acceleration and gravity of the body, using a bandpass filter and a third-order Butterworth lowpass, both with a cutoff frequency of 20 Hz to eliminate noise. Gravity has only low frequency components, therefore a lowpass Butterworth cutoff frequency of 0.3 Hz filter was used. From each window feature vector is obtained by calculating the variables of time and frequency domain.

Body acceleration and angular velocity derived function of time, for the jerk signals, and the magnitude of these signals was calculated dimensional handling the Euclidean norm (distance from the origin).

a Fast Fourier Transform (FFT) it was used in some of the signs that were used to estimate variables vector characteristics, which provided a data matrix of 10,299 samples and 561 features in time domain and frequency (Linchman, 2013).

Metrics

The Database Machine Learning Repository at the University of California has 33 variables obtained from the signals in the three-axis accelerometer and gyroscope, which were processed with 17 metric. This gives a total of 561 features, resulting from the multiplication between variables and metrics. Then metrics and variables corresponding tables are observed.

Table 1 contains some of the same but in different axes variables, why the variable is counted three times for the three axes (X, Y, Z).

Table 1. Set of Variables

#	Descripción
1,2,3	Aceleración del cuerpo en los tres ejes (XYZ), en función del tiempo.
4,5,6	Aceleración de la gravedad en los tres ejes (XYZ), en función del tiempo.
7,8,9	Derivada de la aceleración del cuerpo en los tres ejes (XYZ), en función del tiempo.
10,11,12	Velocidad angular del cuerpo en los tres ejes (XYZ), en función del tiempo.
13,14,15	Derivada de la velocidad angular del cuerpo en los tres ejes (XYZ), en función del tiempo.
16	Magnitud de la aceleración del cuerpo, en función del tiempo.
17	Magnitud de la aceleración de la gravedad, en función del tiempo.
18	Magnitud de la derivada de la aceleración del cuerpo, en función del tiempo.
19	Magnitud de la velocidad angular del cuerpo, en función del tiempo.
20	Magnitud de la derivada de la velocidad angular del cuerpo, en función del tiempo.
21,22,23	Aceleración del cuerpo en los tres ejes (XYZ), en dominio de la frecuencia.
24,25,26	Derivada de la aceleración del cuerpo en los tres ejes (XYZ), en dominio de la frecuencia.
27,28,29	Velocidad angular del cuerpo en los tres ejes (XYZ), en dominio de la frecuencia.
30	Magnitud de la aceleración del cuerpo, en dominio de la frecuencia.
31	Magnitud de la derivada de la aceleración del cuerpo, en dominio del tiempo.
32	Magnitud de la velocidad angular del cuerpo, en dominio de la frecuencia.
33	Magnitud de la derivada de la velocidad angular del cuerpo, en dominio de la frecuencia.

Table 2. Set of Metrics

#	Métricas
1	Media
2	Desviación Estándar
3	Desviación Media Absoluta
4	Valor Máximo
5	Valor Mínimo
6	SMA
7	Energía
8	IQR
9	Entropía
10	Auto regresión
11	Correlación
12	Máximo Índice
13	Frecuencia Media
14	Skewness
15	Kurtosis
16	Energía de un intervalo de frecuencia
17	Ángulo entre vectores

Information Gain

He mentioned that the criterion of maximizing the information gain is based on the entropy of information theory, ie it is a measure of uncertainty of a random variable (Roobaert, Karakoulas, and Chawla, 2006).

To determine the information gain of this study attributes are discretized, from which the gain for the 5 groups ranges as calculated; thereby an ordered list was generated and removed those attributes with lower results.

The 5 groups were used: 4, 6, 8, 10 and 12 intervals; the reason for using only these groups is because intervals show a general tendency, that is, if a greater or lesser number of intervals (less than 4 or more than 12), the average remains the same is used, which does not change the data.

In conclusion, the expected reduction in the entropy of the data to know the value of continuous attributes in classification tasks of physical activity was performed.

Neural networks

Its main function is learning scheme and determines the type of problems that will be able to solve (Isasi and Galvan, 2004). On the other hand, researchers from artificial intelligence and statistics have been interested in the most abstract properties of neural, such as its ability to develop distributed computing and tolerate the noise at the input of the (Cazorla, Colomina Pardo network networks, and Old Hernando, 2011). Currently it is understood that other types of systems (including Bayesian networks) have these properties, however, neural networks are worthy of study because they remain as one of the most popular and effective when building learning systems (Russell and Norving forms, 2004).

In the present work it is to know what type of neural network offers greater efficiency when meet the proposed requirements. It was decided networks feedforward (feedforward), which contain a series of layers: one with input connection to the network, a backing layer that has a connection to the previous layer and a layer that produces the output of the network . These networks with enough neurons in the hidden layer can be adapted to any problem of input-output mapping finite. Two types of feedforward networks known and used in the mathematical tool MATLAB® are FITNET and Patternnet.

FITNET is a feedforward network with two layers sigmoid activation function: a layer of hidden neurons and other neuron linear output, which engage multidimensional problems with consistent data allocation. The network will be trained with backpropagation algorithm "Levenberg-Marquardt", and if the memory is not sufficient propagation algorithm is used back scaled conjugate gradient (Monar, 2014).

Networks pattern recognition (Patternnet) is a feedforward network of two layers: one with a hidden and output neurons Softmax (net pattern) with transfer function sigmoid type. This network is trained with the algorithm backpropagation scaled conjugate gradient for sorting the outputs according to the inputs, and the target data should consist of vectors of all zero values except for a 1 in the (i) which it is the class to represent (Monar, 2014).

RESULTS AND ANALYSIS

In order to select the neural network that offers improved efficiency relative to processing time and error rate, three training trials were performed with 1, 70 and 561 features respectively, as shown in Table 3.

Table 3. Training neural networks.

Red Neuronal	Número de Características	Número de Muestras	Tiempo de Procesamiento (s)	Porcentaje de Error Train (%)	Porcentaje de Error Test (%)
Fitnet	1	7 352	6.296	15.821	15.845
	70	7 352	97.340	1.352	4.515
	561	7 352	325.761	0.4528	1.8709
Patternnet	1	7 352	9.647	29.491	29.583
	70	7 352	44.226	1.990	5.417
	561	7 352	112.262	0.701	2.322

Referring to the results of the training of each network, it is observed that the Patternnet network has less time for data processing and a percentage of considerable error compared to the FITNET network, allowing the status of physical activity a or more people simultaneously and in real time; those reasons support the use of this network.

Subsequently gain characteristic information that each dividing the set of training data into 5 groups containing different numbers of intervals is calculated. These were structured as follows: the first group divided into 4 intervals, the second in 6 intervals, the third in 8 intervals, the fourth in 10 intervals and the fifth in 12 intervals. This was done because the data are continuous attributes (Cao, Ma, Liu and Guo, 2012).

Features accelerometer and gyroscope

The features were ordered from one that provides as much information up to the least contributes, as shown in Figure 1.

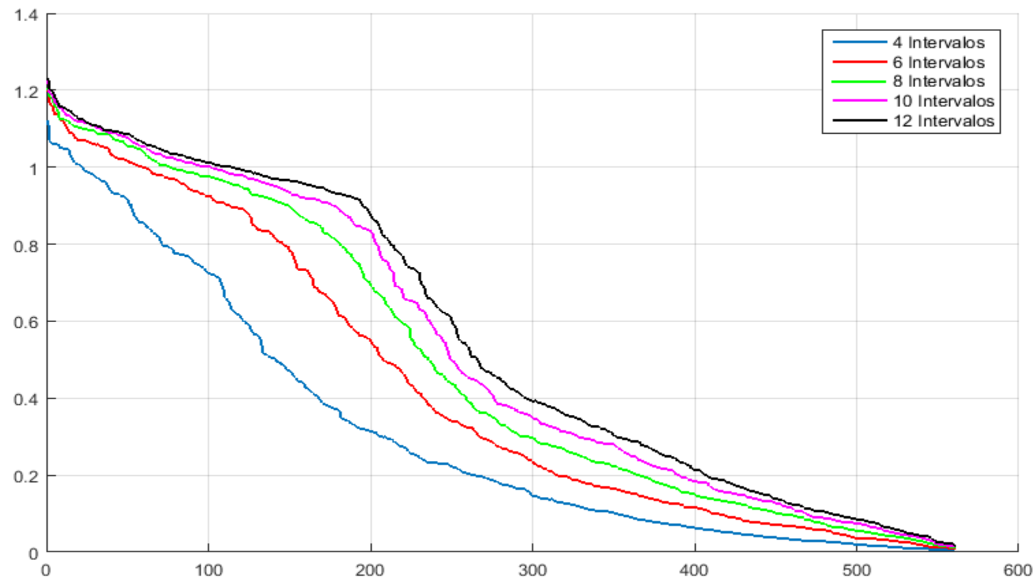

Figure 1.- Gain information for 5 groups of intervals (561 features).

When calculating the derivative gain information you can appreciate the break points in each group of intervals, for example, about 200 features is clearly the fall of the curve from this the mean (average) is obtained, which later serve to set the number of features and the respective error mark both training (train) and testing (test), which can be seen in table 4.

Table 4. Selecting break points.

	INTERVALOS						INTERVALOS					
	4	6	8	10	12	Promedio	4	6	8	10	12	Promedio
NÚMERO DE CARACTERÍSTICAS	280	276	278	280	278	278	133	131	139	129	133	133
	267	266	263	263	254	263	127	123	125	125	126	125
	261	259	259	254	253	257	103	102	104	101	109	104
	250	248	252	247	248	249	85	94	93	86	89	89
	234	234	232	231	229	232	85	83	85	81	78	82
	230	229	224	227	229	228	72	71	69	72	70	71
	211	214	213	215	213	213	61	63	60	64	61	62
	206	205	208	204	203	205	51	52	47	50	50	50
	198	201	195	192	191	195	34	43	40	38	39	39
	193	188	184	185	181	186	24	25	30	25	22	25
	178	179	170	179	171	175	16	14	13	16	14	15
	171	169	161	164	168	167	7	7	7	7	7	7
	157	150	150	152	151	152	5	4	4	4	3	4
	137	142	142	140	141	140	1	1	1	1	1	1

Through the application all the attributes identical characteristics were determined in the 5 groups of intervals, identifying that which way are contributing more information relevant to the process of classification of physical activity for the specific case.

Clearly the error rate for both the training set and for the test, is gradually increased while the characteristics are reduced. Thus the number of minimum data capable of generating an error less than 5% for the neural network classifier can learn and detect patterns that determine physical activity being performed is chosen.

Table 5. Reduction accelerometer and gyroscope features with identical characteristics.

Número Características	Número Características Idénticas	Train (%)	Test (%)	Número Características	Número Características Idénticas	Train (%)	Test (%)
561	561	0.7743	2,9446	133	133	1.6391	3.6263
278	235	1.4187	3,1113	125	125	1.7344	3.6958
263	224	1.587	3,2117	104	104	1.6841	3.8639
257	221	1.4117	2,9988	89	89	6.6016	7.0473
249	218	1.5557	3,1409	82	82	6.9032	7.3374
232	213	1.4177	3,1252	71	71	7.0418	7.6495
228	211	1.5069	3,0083	62	62	7.204	7.6347
213	205	1.738	3,3246	50	50	7.3649	8.0591
205	201	1.547	3,0004	39	39	8.8933	8.7641
195	194	1.5908	3,2094	25	25	10.197	10.2485
186	185	1.3913	3,1513	15	15	12.2387	13.3877
175	174	1.4069	3,1168	7	7	13.216	14.8047
167	167	1.4222	3,0931	4	4	13.2696	14.865
152	152	1.5395	3,3428	1	1	15.2551	15.9221
140	140	1.6762	3.5242				

As evidenced in Table 5, with decreasing the number of features tends to be compared with the number of identical characteristics. Thus, if observed for the number of features 167, it has the same number of number of identical characteristics, and these characteristics diminish as the error tends to increase. Therefore, by making the combined calculation for the accelerometer to the gyroscope a joint decreased to 104 characteristics is obtained, with an error of test assembly 3.8639%, it becomes the limit of decreasing characteristics before passing 5% error.

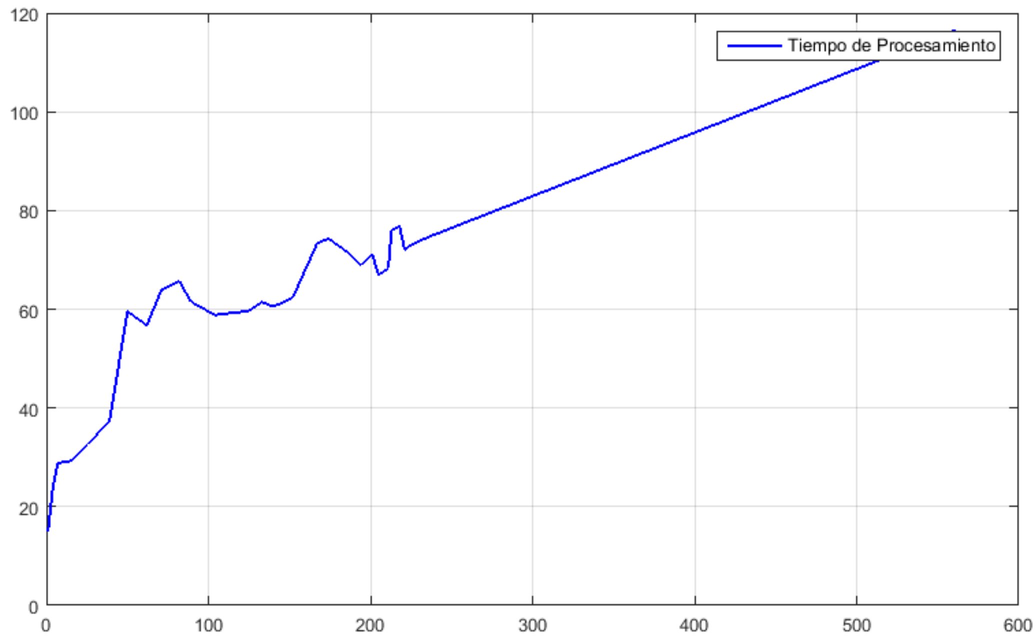

Figure 2.- Processing time (all features).

By decreasing the number of features as shown in Figure 2, shows an efficient reduction of processing time reaching optimal results of 58.8267 (s) 104 for characteristics.

Features Accelerometer

The database used provides features 561, 345 of which are accelerometer and gyroscope 216.

Once this data analysis classified the same selection and attribute reduction is conducted, taking the characteristics of the accelerometer and gyroscope separately as shown in the following tables and figures.

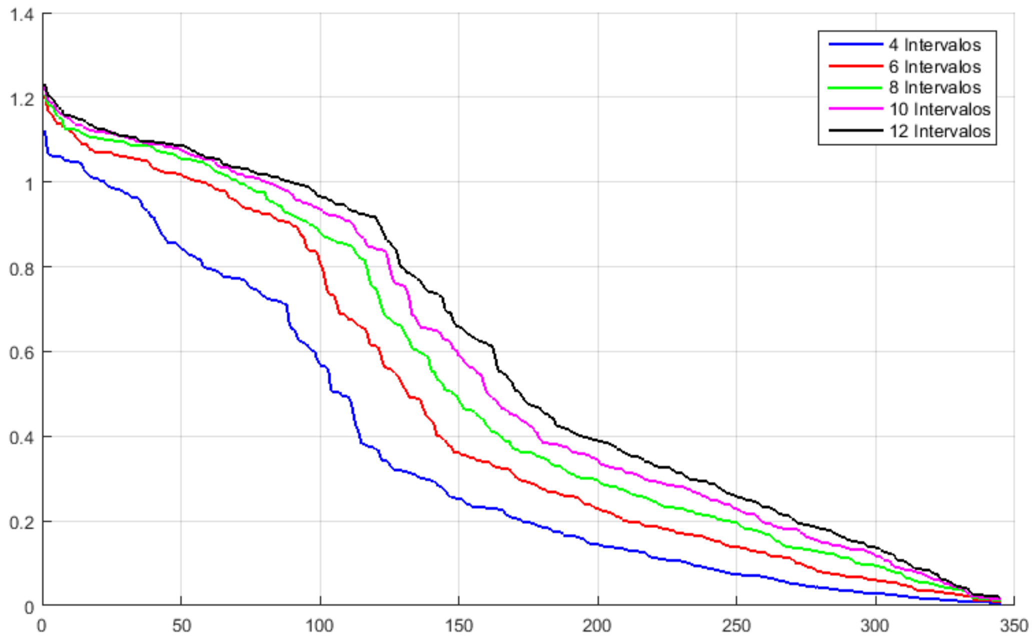

Figure 3.- Gain information for 5 groups of intervals (345 features accelerometer).

Through the derivative gain information you can appreciate the break points in each group interval. In the case of accelerometer specifically a drop characteristics curve at approximately 120, as shown in Figure 3. Table 6 is set to the average number of features set with which the respective error calculation marking will be made is denoted both the training set and the test.

Table 6. Selecting break points (accelerometer).

	INTERVALOS						INTERVALOS					
	4	**6**	**8**	**10**	**12**	**Promedio**	**4**	**6**	**8**	**10**	**12**	**Promedio**
NÚMERO DE CARACTERÍSTICAS	195	194	200	200	206	199	88	92	86	84	85	87
	187	181	182	178	184	182	78	82	80	78	74	78
	166	168	168	167	169	168	73	70	70	67	67	69
	152	161	159	159	163	159	64	61	61	62	64	62
	146	147	149	147	147	147	49	50	47	43	46	47
	144	141	144	143	144	143	38	38	40	33	34	37
	125	129	132	132	128	129	22	25	30	25	24	25
	121	122	121	124	122	122	14	16	13	14	16	15
	103	100	99	100	98	100	7	7	7	7	7	7
	91	94	95	90	96	93	1	1	1	1	1	1

From points calculated break, shown in Table 7 that adequate decrease before exceeding 5% error, corresponds to 78 characteristics, demonstrating that independently a greater decrease in characteristics is obtained in the case of the accelerometer . It should be noted that applies the same calculation and analysis for the case of the gyroscope.

Table 7. Accelerometer reduction characteristics with identical characteristics.

Número Características	Número Características Idénticas	Train (%)	Test (%)	Número Características	Número Características Idénticas	Train (%)	Test (%)
345	345	0.7721	3.0793	87	87	1.3014	4.1031
199	164	1.0944	3.3315	78	78	2.9037	4.1573
182	155	1.1282	3.244	69	69	6.8814	8.3506
168	148	1.1226	3.5088	62	62	7.3665	8.47
159	145	1.1337	3.4383	47	47	8.1803	9.1972
147	140	1.279	3.3502	37	37	7.5788	9.337
143	138	1.1135	3.2756	25	25	9.6996	11.576
129	129	1.3649	3.6542	15	15	11.1481	12.1537
122	122	1.0565	3.8856	7	7	12.7406	13.3385
100	100	1.4052	4.0535	1	1	15.228	15.8485
93	93	1.1843	3.9107				

Similarly it can be seen in Figure 4 that the 78 identical characteristics are processed in 55.9645 (s). This generates a significant time reduction with respect to the processing of data from both sensors.

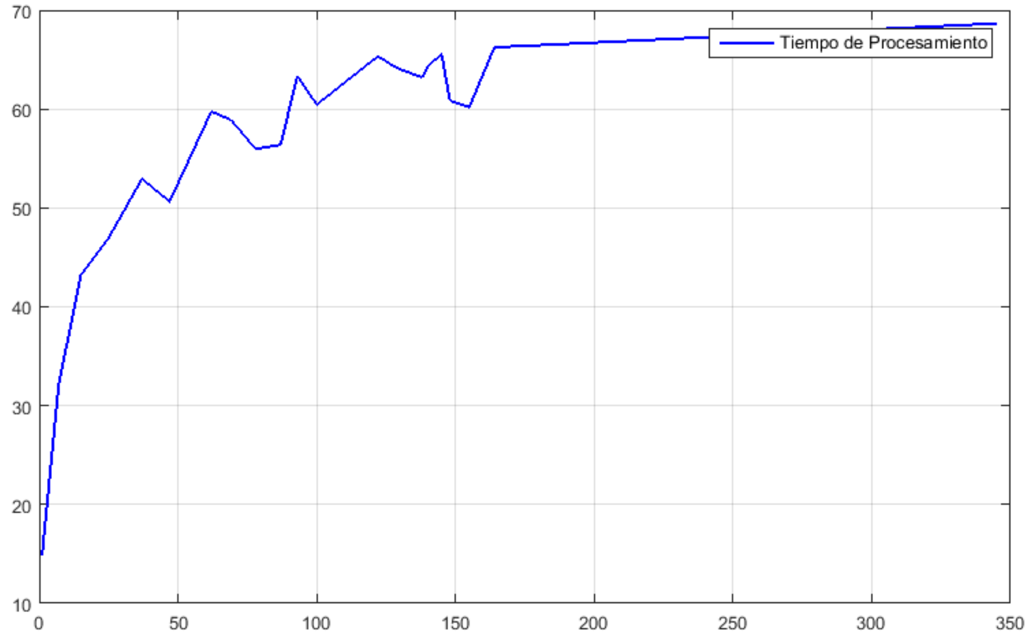

Figure 4.- Processing time (accelerometer features).

Features Gyroscope

Figure 5 shows a gain of different information for the 5 groups of intervals, so that in calculating its derivative the break points will be more apparent. In this case a drop of the curve around 80 characteristics is observed. Consecutively selection and analysis only attribute reduction characteristics gyroscope is made.

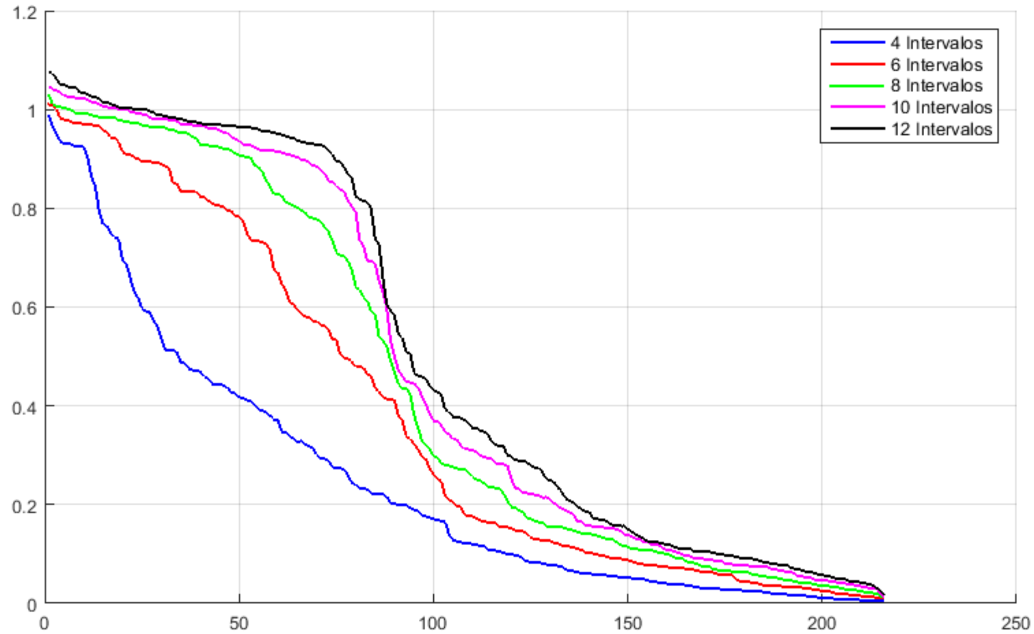

Figure 5.- Gain information for 5 groups of intervals (216 features gyroscope).

Table 8 shows an obvious decrease in the number of break points compared with previous results, just as the mean number of characteristics is determined and then the error rate for the training set and test is calculated.

Table 8. Selecting break points (gyroscope).

	INTERVALOS					
	4	6	8	10	12	Promedio
NÚMERO DE CARACTERÍSTICAS	133	138	147	136	140	139
	122	124	128	122	128	125
	113	116	118	119	118	117
	103	102	101	102	102	102
	96	92	94	96	94	94
	88	84	85	88	86	86
	78	75	79	80	79	78
	60	58	61	53	55	57
	47	51	48	47	46	48
	34	32	39	34	35	35
	19	19	18	21	18	19
	13	16	12	14	14	14
	1	**3**	1	4	**3**	2

The fewer points calculated break, as judged in Table 9, it shows that proper decrease before exceeding 5% error, corresponds to 78 characteristics, demonstrating that independently characteristics gyroscope can also be used for the classifier.

Table 9. Reduction gyroscope features with identical characteristics.

Número Características	Número Características Idénticas	Train (%)	Test (%)
216	216	0.758	3.1706
139	116	1.257	4.1913
125	110	1.2988	4.2884
117	109	1.299	4.2975
102	102	1.4572	4.2413
94	94	2.025	3.7185
86	86	1.9113	4.0126
78	78	2.0773	3.7174
57	57	5.8204	6.3557
48	48	6.3161	6.6827
35	35	7.4344	7.5665
19	19	9.6658	11.5729
14	14	10.1099	12.0147
2	2	16.0832	17.1085

Similarly, the graphics processing time for the characteristics expressed gyroscope efficient reduction equal to the accelerometer values, as shown in Figure 6.

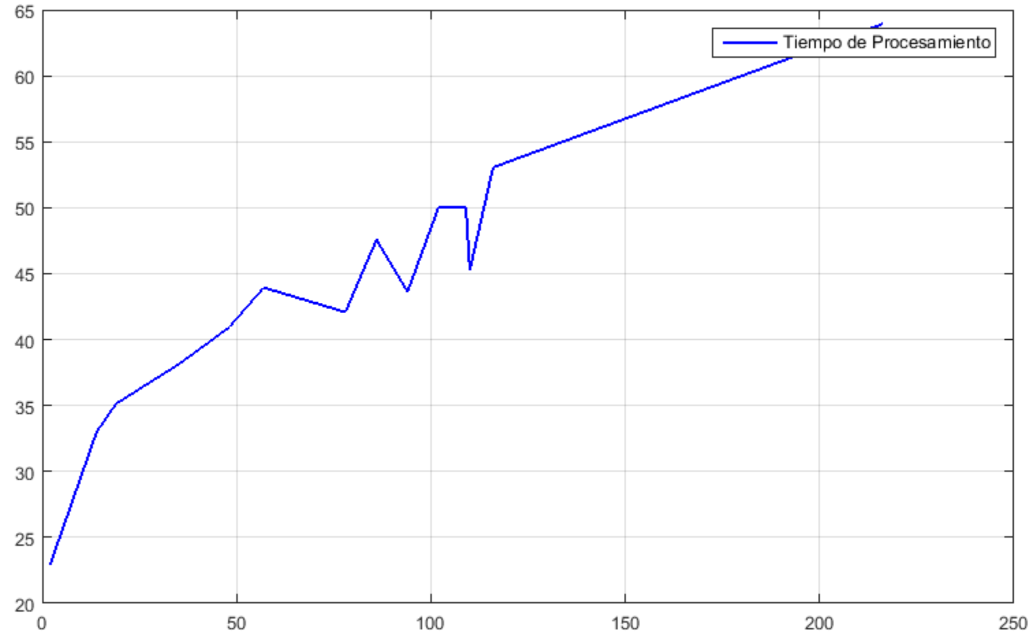

Figure 6.- Processing Time (gyroscope features).

In the case of the accelerometer, in about 78 identical characteristics permissible error range (3.7185%) is obtained before exceeding an error of 5% and time efficient processing of 42. 0811 seconds.

The analysis of the characteristics for each sensor checks that the selection process is more efficient if working independently. While the mobile device has two sensors (accelerometer and gyroscope), it is preferable that the classification process characteristics is made with one of the two, so be achieved optimize the processing time. This result is not achieved efficiently if both sensors work in tandem.

Conclusion

14% of continuous attributes sampled is used, and it is determined that one can obtain a reduction in processing time by about 55% and an error less than 5% in the selection process characteristics without affecting the classification of activity physical.

In the course of the investigation it was established by reducing features for continuous attributes, which can meet the physical activity and / or status of a person (walking, jumping, running, etc.) so more efficiently by Gyroscope or accelerometer independently.

This article serves as a foundation for future work with the approach of other methods. One can cite as examples the improvement information gain algorithm through the introduction of the dependence of attributes. So, you get to collect data not only from one person, but several simultaneously and processing information in real time.

Bibliography

Cao, D., Ma, N., Liu, Y., & Guo, J. (2012). A Feature Selection Algorithm for Continuous Attributes Based on the Information Entropy. *Journal of Computational Information Systems*, 1467-1475.

Cazorla, M., Colomina Pardo, O., y Viejo Hernando, D. (19 de mayo de 2011). Presentaciones de la asignatura Técnicas de Inteligencia Artificial (Curso 2010-2011). Obtenido de http://hdl.handle.net/10045/17323

Das, S., Green, L., Perez, B., & Murphy, M. (30 de julio de 2010). Detecting User Activities using the Accelerometer on Android Smartphones. *TRUST REU The Team for Research in Ubiquitous Secure Technology*, 29.

Han, J., Kamber, M., & Pei, J. (2011). *Data Mining: Concepts and Techniques,* tercera edición, USA: Elsevier.

Hong, S. J. (1997). Use of contextual information for feature ranking and discretization. *IEEE transactions on knowledge and data engineering, 9*(5), 718-730.

Isasi, P., y Galván, I. (2004). *Redes De Neuronas Artificiales. Un enfoque práctico*, primera edición, Madrid, España: Pearson.

Kwapisz, J., Weiss, G., & Moore, S. (2011). *Activity recognition using cell phones accelerometers. ACM SIGKDD Explorations Newsletter*, segunda edición, vol. 12, New York, USA.

Linchman, M. (04 de abril de 2013). UCI Machine Learning Repository. Recuperado el 10 de noviembre de 2015, de http://archive.ics.uci.edu/ml

Lu, H., Setiono, R., & Liu, H. (1996). Effective data mining using neural networks. *IEEE transactions on knowledge and data engineering, 8*(6), 957-961.

Mitchell, E., Monaghan, D., & O'Connor, N. (19 de abril de 2013). Classification of Sporting Activities Using Smartphone Accelerometers. *Sensors, 13*, 16.

Monar, W. L. (octubre de 2014). Repositorio Digital - Escuela Pilitécnica Nacional. Recuperado el 30 de julio de 2016, de http://bibdigital.epn.edu.ec/handle/15000/8711

Roobaert, D., Karakoulas, G., & Chawla, N. (2006). Information gain, correlation and support vector machines. *Feature Extraction. Springer Berlin Heidelberg, 207*, 463-470.

Russell, S., y Norving, P. (2004). *Inteligencia Artificial un enfoque Moderno,* segunda edición, Madrid, España: Pearson.

Weiss, G., & Hirsh, H. (27 de agosto de 1998). Learning to Predict Rare Events in Event Sequences. *Knowledge Discovery and Data Mining*.

Yang , B., & Wang, L. (2011). *The Construction Method of Knowledge Discovery Theory System Based on Cognitive.* (C. a. Circuits, Ed.) Wuhan: IEEE.