

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

*Feature selection for continuous attributes in physical activity classification 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](mailto:evcarrera@espe.edu.ec)

**Jefferson Stalin Rodríguez Páramo**

Universidad de las Fuerzas Armadas ESPE, Ecuador

[jsrodriguez2@espe.edu.ec](mailto: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 include many sensors capable of sending data used in decision making activities; an example is the classification of physical activity based on the use of accelerometers and gyroscopes. Sensor's signals could be processed previously applying different techniques that extract many attributes that would be used in the development of classification tasks. The optimization of the classification system required the reduction of the number characteristics in order to synthesize the set dimension and the learning time. This article uses metrics of information gain based on continuous attributes, these metrics reduced the uncertainty and extracted only the most significant across the processed information. The analysis of the results obtained in the classification of physical activity using neural networks showed not only a decrement on the number of characteristics, but also, an error less than 5 % and reducing processing time in approximately 55 %.

**Key words:** machine learning, physical activity, feature 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 atividade física baseada na utilização de acelerômetros 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 otimização 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 atividade 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

---

## Introdução

Os dispositivos móveis conter vários sensores que são actualmente utilizados em vários campos e inúmeras aplicações em todo o mundo (Das, Verde, Perez, e Murphy, 2010). Os "inteligentes" dispositivos móveis devido ao seu pequeno tamanho, capacidade de enviar e receber dados e poder de computação, pode armazenar informações que podem ser manipulados (Kwapisz, Weiss, e Moore, 2011).

Todas as informações coletadas por esses dispositivos electrónicos fornece uma contribuição significativa para o desenvolvimento e monitoramento, aspectos relacionados aos cuidados de saúde, reabilitação, diagnóstico da doença, a segurança de pessoas, entre outros (Mitchell, Monaghan, e O'Connor, 2013 ).

Os sinais que são emitidos pelos sensores não pode ser classificado com algoritmos padrão, de modo que, em primeira instância devem ser transformados dados do estado de pura informação, a transformação é a função mais simples de tempo ou frequência (Weiss e Hirsh, 1998). Assim, o processamento é conseguido e extraiu-se uma série de recursos com base em diferentes métricas.

A grande quantidade de dados a partir da entrada em vigor faz com que o tempo de processamento aumenta (Han, Kamber, e Pei, 2011), o que faz com que a otimização do sistema de classificação de redução dessas exigências. Isto exige a utilização de um algoritmo que

permite a selecção de características, de modo que a dimensão de tempo em conjunto e aprendizagem (Yang e Wang, 2011) é sintetizado.

A classificação ou automático seleção de recursos é uma das tarefas mais comuns onde as redes neurais artificiais têm se mostrado eficazes, como eles executam um processamento automático de dados e baseiam-se no sistema nervoso biológico (Isasi e Galvan, 2004).

É digno de nota que a Redes Neurais Artificiais, de sua aparência e seu rápido desenvolvimento têm tido um uso significativo como uma tecnologia para a mineração de dados, porque a tecnologia tem atributos para modelagem eficaz e eficiente de problemas complexos (Lu, Setiono, e Liu, 1996).

Esta pesquisa é baseada em processamento de informação do tipo contínuo usando métricas para obter informações, que por meio de um algoritmo será quantificada. Assim, o processo de seleção de recursos para os atributos contínuos em tarefas de classificação de atividade física será possível; ou seja, as características mais importantes para o processo de classificação serão identificados, ao reduzir a incerteza e obter apenas os mais significativos.

Características seleccionadas deve especificar a actividade física de uma pessoa (subir ou descer escadas, sentado, em pé, deitado).

Vale ressaltar que o critério de maximizar o ganho de informação produz um viés para os atributos que têm lotes de diferentes valores, que resolve esse problema usando a relação do lucro como critério de separação (Hong, 1997). Esta medida leva em conta tanto o ganho de informação como as probabilidades dos diferentes valores dos atributos; por sua vez, estas probabilidades são recolhidos por a informação de separação, o que não é mais do que a entropia do conjunto de dados a partir dos valores dos atributos.

Os resultados da classificação da actividade física utilizando redes neurais, como descrito mais adiante, mostra que, ao utilizar: ganho de informação, pontos de quebra para cinco grupos de

intervalos de seleção e taxa de erro em cada, foi conseguida diminuir o conjunto de recursos (561) em 86% (78), de modo a otimização é percebida no processamento de dados em tempo.

A estrutura deste artigo é a seguinte: na Seção 1 da descrição de materiais e métodos adoptados é mostrado, que expõem desenvolvimento experimental para a captura de dados, a descrição matemática do algoritmo proposto eo processo de formação rede. Na seção 2 os resultados experimentais e sua análise são exibidos; e, finalmente, na seção 3 das conclusões.

## **1. MATERIAIS E MÉTODOS**

### **Conjunto de dados e sensores**

No mercado há uma grande variedade de dispositivos móveis que têm sido desenvolvidos diferentes sistemas operacionais, como o iOS da Apple eo Android do Google. banco de dados "Reconhecimento de Atividade Humana Usando Smartphones conjunto de dados" (UCI HAR Dataset) Repositório de Machine Learning na Universidade da Califórnia, mesmo que trabalhar com um Smartphone (Samsung Galaxy S II) foi utilizada neste trabalho colocado na cintura. Através do seu acelerômetro e giroscópio embutido, aceleração linear e sua velocidade angular em três eixos XYZ é obtido. Os experimentos foram filmados para rotular os dados manualmente. O conjunto de dados é dividida aleatoriamente em dois grupos. Lá, 70% (21 pessoas) de voluntários foi selecionado para gerar os dados de treinamento e 30% (9 pessoas) dados de teste fornecidos.

Selecionados para este características de banco de dados derivados dos sinais brutos do acelerómetro de três eixos e giroscópio, que no domínio do tempo foram capturadas a uma velocidade constante de 50 Hertz (Hz) e foram amostrados com amplas janelas de correr 2,56 segundos fixos (s) e 50% de sobreposição (128 leituras / janela). O sinal de aceleração tem dois componentes: o movimento gravitacional e do corpo; que eles são separados e refinado em aceleração e gravidade do corpo, utilizando um filtro passa-banda e um de terceira ordem Butterworth passa baixo, ambas com uma frequência de corte de 20 Hz para eliminar o ruído. A gravidade tem apenas componentes de baixa frequência, portanto, foi utilizado um lowpass

Butterworth frequência de corte de 0,3 filtro Hz. De cada vector recurso janela é obtido através do cálculo das variáveis de tempo e domínio da frequência.

aceleração do corpo e da velocidade angular função do tempo derivada, para os sinais de empurrão, e a magnitude destes sinais foi calculada dimensional manusear a norma euclidiana (distância a partir da origem).

uma Transformada Rápida de Fourier (FFT) foi utilizado em alguns dos sinais que foram utilizados para estimar as características variáveis vector, o que proporcionou uma matriz de dados de 10,299 e 561 amostras de características no domínio do tempo e frequência (Linchman, 2013).

### **Métrica**

O banco de dados repositório Aprendizado de Máquinas da Universidade da Califórnia tem 33 variáveis obtidas a partir dos sinais do acelerômetro de três eixos e giroscópio, que foram processados com 17 métrica. Isto dá um total de 561 recursos, resultante da multiplicação entre as variáveis e métricas. Em seguida, métricas e variáveis tabelas correspondentes são observados.

A Tabela 1 contém alguns dos mesmos, mas em diferentes variáveis machados, porque a variável é contada três vezes para os três eixos (X, Y, Z).

Tabla 1. Conjunto de 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.

Tabla 2. Conjunto de Métricas

#	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

## **Ganho de informação**

Ele mencionou que o critério de maximizar o ganho de informação é baseado na entropia da teoria da informação, ou seja, é uma medida da incerteza de uma variável aleatória (Roobaert, Karakoulas, e Chawla, 2006).

Para determinar o ganho de informação deste estudo são atributos discretizado, a partir do qual o ganho para os 5 grupos varia conforme calculado; assim, uma lista ordenada foi gerado e removeu esses atributos com resultados mais baixos.

Utilizaram-se os grupos de 5: 4, 6, 8, intervalos de 10 e 12; a razão para a utilização de apenas estes grupos é porque intervalos de mostrarem uma tendência geral, isto é, se um número maior ou menor de intervalos (menos de 4 ou mais de 12), a média, mantém-se o mesmo é utilizado, o qual não faz alterar os dados.

Em conclusão, a redução esperada na entropia dos dados para saber o valor de atributos contínuos em tarefas de classificação de atividade física foi realizada.

## **Redes neurais**

Sua principal função é aprender sistema e determina o tipo de problemas que serão capazes de resolver (Isasi e Galvan, 2004). Por outro lado, pesquisadores de inteligência artificial e as estatísticas têm sido interessado nas propriedades mais abstratas de neural, tais como a sua capacidade de desenvolver computação distribuída e tolerar o barulho na entrada do (Cazorla, rede redes Colomina Pardo e Velha Hernando, 2011). Atualmente entende-se que outros tipos de sistemas (incluindo redes Bayesian) têm essas propriedades, no entanto, as redes neurais são dignos de estudo, pois eles permanecem como um dos na construção de sistemas de aprendizagem (formas Russell e Norving mais populares e eficazes, 2004).

No presente trabalho é saber que tipo de rede neural oferece maior eficiência ao atender os requisitos propostos. Foi decidido redes feedforward (alimentação de entrada), que contém uma



série de camadas: uma ligação de entrada com a rede, uma camada de suporte que tem uma conexão para a camada anterior e uma camada que produz a saída da rede . Estas redes de neurónios com suficiente na camada escondida pode ser adaptado a qualquer problema de mapeamento finito de entrada-saída. Dois tipos de redes feedforward conhecidos e utilizados na MATLAB ferramenta matemática são FITNET e Patternnet.

FITNET é uma rede feedforward com função de ativação sigmóide duas camadas: uma camada de neurónios ocultos e outros saída do neurônio linear, que envolvem problemas multidimensionais com alocação de dados consistente. A rede será treinado com o algoritmo backpropagation "Levenberg-Marquardt", e se a memória não é suficiente algoritmo de propagação é usado gradiente conjugado volta dimensionado (Monar, 2014).

reconhecimento de padrões Networks (Patternnet) é uma rede feedforward de duas camadas: uma com um ocultos e saída neurónios softmax (padrão de rede) com o tipo de função de transferência sigmóide. Esta rede é treinada com o algoritmo backpropagation escalado gradiente conjugado para classificar as saídas de acordo com as contribuições, e os dados de destino deve ser composto de vetores de todos os valores zero, exceto para um 1 na (i) que é a classe para representar (Monar, 2014).

## **RESULTADOS E ANÁLISE**

De modo a seleccionar a rede neural que oferece uma melhor eficiência em relação ao tempo de processamento e taxa de erro, três ensaios de formação foram realizadas com 1, 70 e 561, respectivamente, características, como se mostra na Tabela 3.

Tabla 3. Entrenamiento de redes neuronales.

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

Referindo-se aos resultados da formação de cada rede, observa-se que a rede Patternnet tem menos tempo para o processamento de dados e uma percentagem de erro considerável em comparação com a rede FITNET, permitindo que o estado físico de uma actividade ou mais pessoas simultaneamente e em tempo real; estas razões apoiar a utilização desta rede.

Posteriormente obter informações característica que cada divisão do conjunto de dados de treinamento em 5 grupos contendo diferentes números de intervalos é calculado. Estes foram estruturadas da seguinte forma: o primeiro grupo dividido em intervalos de 4, o segundo em intervalos de 6, o terceiro em 8 intervalos, a quarta em intervalos de 10 eo quinto em 12 intervalos. Isso foi feito porque os dados são atributos contínuos (Cao, Ma, Liu e Guo, 2012).

### **Características acelerômetro e giroscópio**

As características foram encomendados a partir de uma que fornece o máximo de informação até às menos contribui, como mostrado na Figura 1.

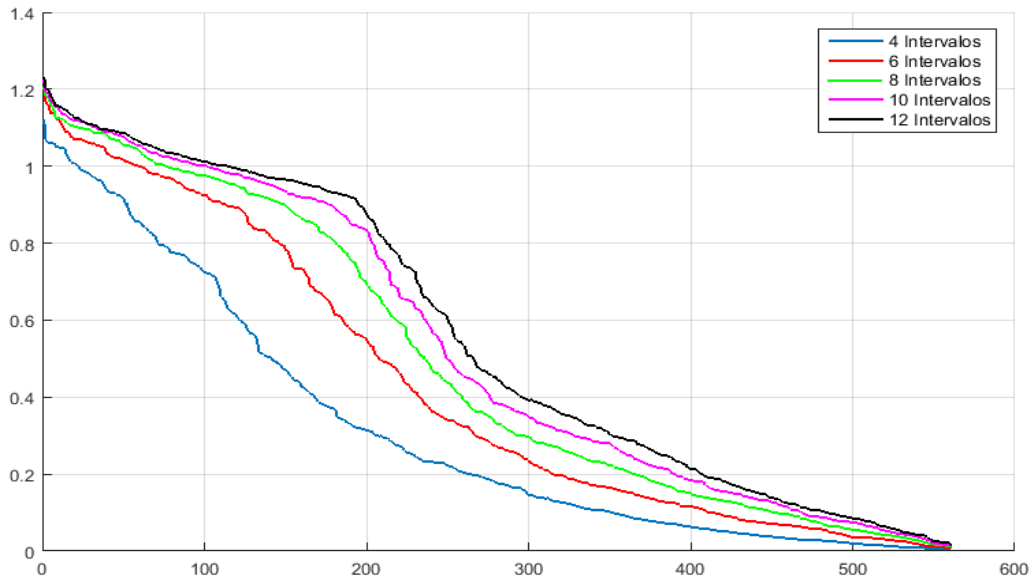


Figura 1.- Ganancia de información para los 5 grupos de intervalos (561 características).

Ao calcular a informação de ganho derivado pode-se apreciar os pontos de quebra em cada grupo de intervalos, por exemplo, cerca de 200 características é claramente a queda da curva a partir deste a média (média) é obtido, o qual mais tarde servem para definir o número de características e o respectivo ponto de erro tanto de formação (comboio) e teste (teste), o que pode ser visto na tabela 4.

Tabla 4. Selección de puntos de quiebre.

	INTERVALOS					Promedio	INTERVALOS					Promedio
	4	6	8	10	12		4	6	8	10	12	
<b>NÚMERO DE CARACTERÍSTICAS</b>	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

Através da aplicação de todos os atributos características idênticas foram determinados em 5 grupos de intervalos, identificando que o caminho que estão contribuindo mais informações relevantes para o processo de classificação da atividade física para o caso específico.

É evidente que a taxa de erro tanto para o conjunto de treino e para o teste, é gradualmente aumentada enquanto que as características são reduzidos. Assim, o número de dados mínima capaz de gerar um erro inferior a 5% para o classificador da rede neural pode aprender e detectar padrões que determinam a actividade física que está sendo executada é escolhido.

Tabla 5. Reducción de características del acelerómetro y giroscopio con características idénticas.

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				

Como evidenciado na Tabela 5, com a diminuição do número de funcionalidades tende a ser comparado com o número de características idênticas. Assim, se observado para a série de características 167, que tem o mesmo número de número de características idênticas, e estas características diminuir à medida que o erro tende a aumentar. Por isso, fazendo o cálculo combinado para o acelerómetro para o giroscópio uma articulação diminuiu para 104

características é obtida, com um erro de montagem de teste 3,8639%, torna-se o limite de diminuir características antes de passar 5% de erro.

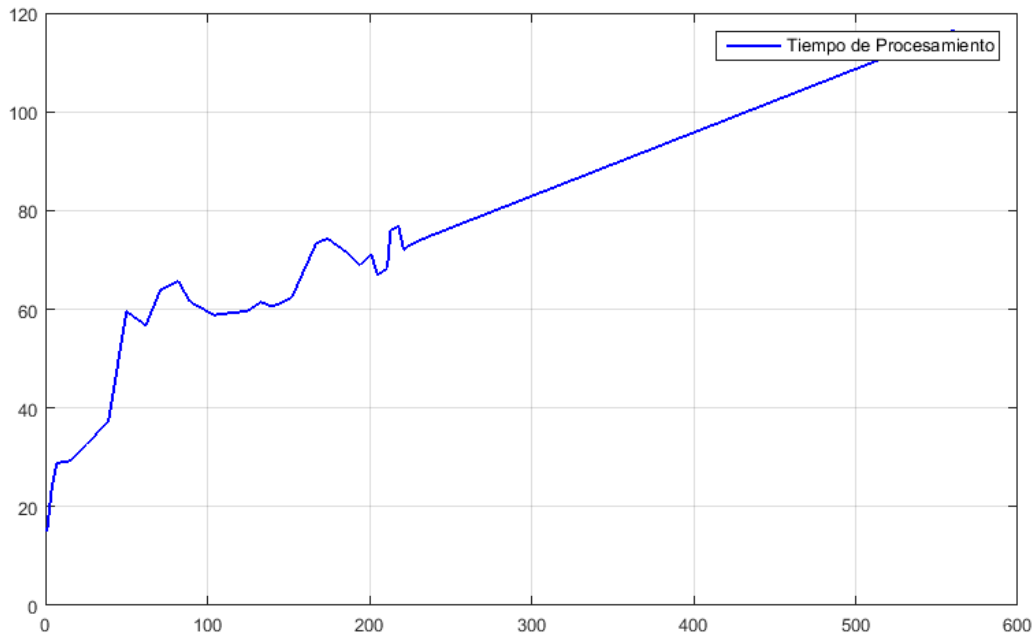


Figura 2.- Tiempo de procesamiento (todas las características).

Ao reduzir o número de recursos, como mostrado na Figura 2, mostra uma redução eficaz do tempo de processamento de atingir resultados óptimos de 58,8267 (s) 104 para as características.

### Características do acelerómetro

O banco de dados utilizado oferece recursos 561, 345 dos quais são acelerómetro e giroscópio 216.

Uma vez que esta análise dos dados classificados e a mesma selecção atributo redução é conduzida, tendo as características do acelerómetro e giroscópio em separado, como mostrado nas seguintes tabelas e figuras.

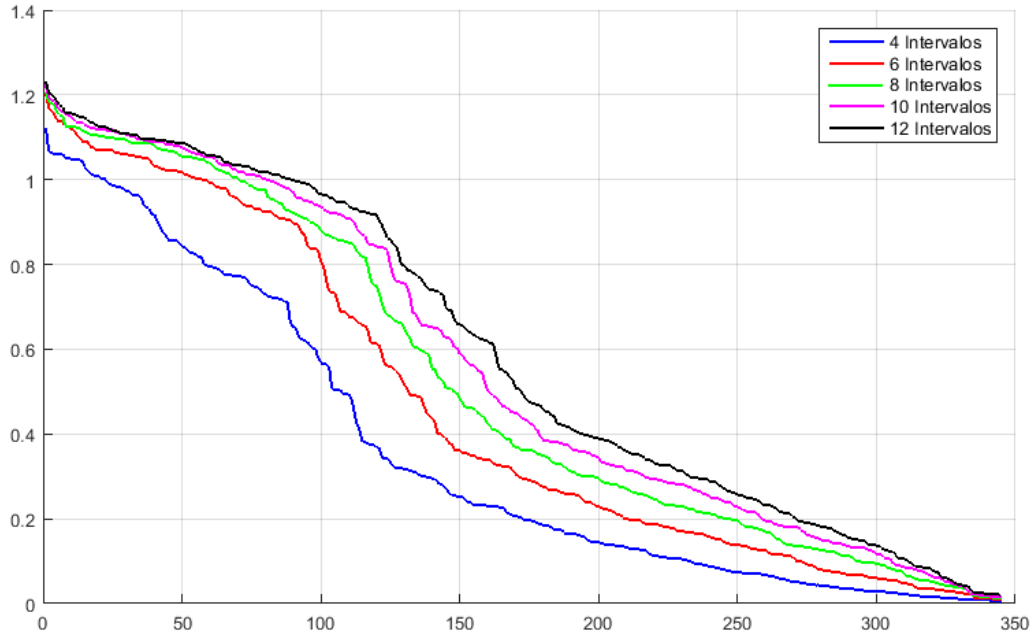


Figura 3.- Ganancia de información para los 5 grupos de intervalos (345 características del acelerómetro).

Através da informação ganho derivativo você pode apreciar os pontos de quebra em cada intervalo de grupo. No caso de uma curva de acelerómetro especificamente características gota a aproximadamente 120, tal como mostrado na Figura 3. Tabela 6 é definido como o número médio de funcionalidades definidas com o qual a respectiva marcação será feita cálculo de erro é denotado tanto o conjunto de treinamento e teste.

Tabla 6. Selección de puntos de quiebre (acelerómetro).

	INTERVALOS					Promedio	INTERVALOS					Promedio
	4	6	8	10	12		4	6	8	10	12	
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

A partir de puntos calculados ruptura, presentados na Tabela 7 que o decréscimo adequado antes de exceder 5% de erro, corresponde a 78 características, o que demonstra que, independentemente, um maior decréscimo na características é obtida, no caso do acelerómetro . Deve notar-se que se aplica o mesmo cálculo e de análise para o caso do giroscópio.

Tabla 7. Reducción de características del acelerómetro con características idénticas.

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				

Do mesmo modo, pode ser visto na Figura 4, que as 78 características idênticas são processados em 55,9645 (s). Isto gera uma redução significativa de tempo no que diz respeito ao processamento de dados de ambos os sensores.

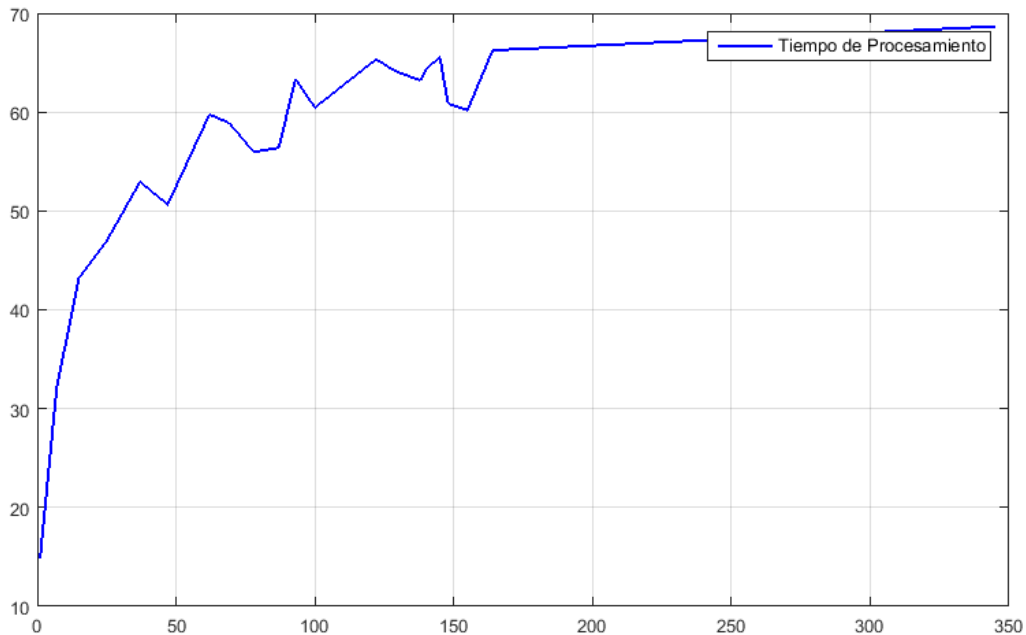


Figura 4.- Tiempo de procesamiento (características del acelerómetro).

### Características giroscópio

A Figura 5 mostra um ganho de informações diferente para os 5 grupos de intervalos, de modo que, no cálculo do seu derivado os pontos de quebra irá ser mais aparente. Neste caso observa-se uma queda da curva 80 em torno de características. Consecutivamente seleção e análise só atribuir características de redução de giroscópio é feita.



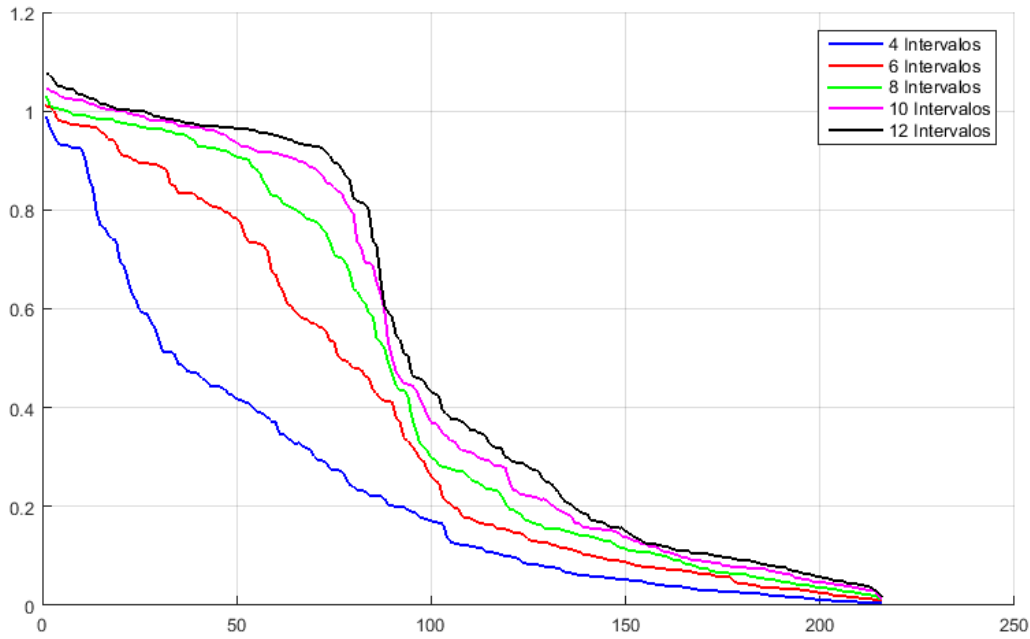


Figura 5.- Ganancia de información para los 5 grupos de intervalos (216 características del giroscopio).

A Tabela 8 mostra uma redução óbvia do número de pontos de quebra em comparação com os resultados anteriores, assim como o número médio de características é determinada e, em seguida, a taxa de erro para o conjunto de treino e de teste é calculado.

Tabla 8. Selección de puntos de quiebre (giroscopio).

	INTERVALOS					Promedio
	4	6	8	10	12	
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

Os poucos pontos de ruptura calculados, como avaliado na Tabela 9, que mostra que o decréscimo adequado antes de exceder 5% de erro, corresponde a 78 características, o que demonstra que independentemente características giroscópio também pode ser utilizada para o classificador.

Tabla 9. Reducción de características del giroscopio con características idénticas.

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

Da mesma forma, o tempo de processamento gráfico para as características expressas redução eficiente giroscópio iguais aos valores do acelerómetro, como mostrado na Figura 6.

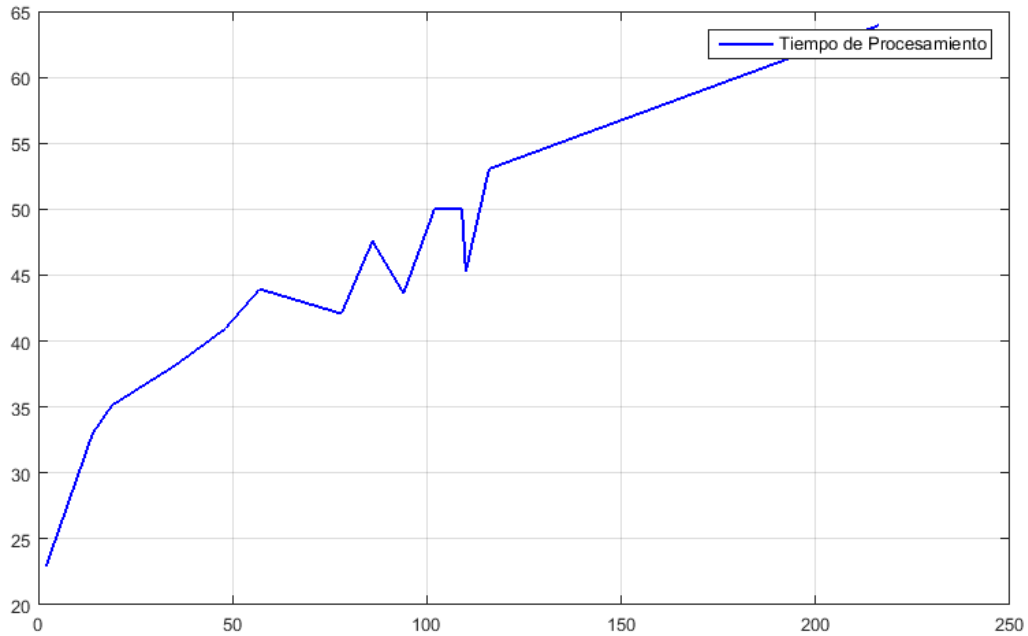


Figura 6.- Tiempo de procesamiento (características del giroscopio).

No caso de o acelerómetro, em cerca de 78 características idênticas gama permissível de erro (3,7185%) é obtido antes de exceder um erro de 5% e de processamento eficiente de tempo 42.0811 segundo.

A análise das características de cada sensor verifica que o processo de selecção é mais eficiente se a trabalhar independentemente. Enquanto o dispositivo móvel tem dois sensores (acelerómetro e giroscópio), é preferível que as características do processo de classificação é feita com um dos dois, de modo a ser conseguida otimizar o tempo de processamento. Este resultado não é conseguida de forma eficiente, se ambos os sensores trabalhar em conjunto.

## **Conclusão**

14% de atributos contínuos amostra é utilizado, e se determinar que se pode obter uma redução no tempo de processamento de cerca de 55% e um erro inferior a 5% nas características do processo de selecção, sem afectar a classificação de actividade física.

No decurso do inquérito, foi estabelecido, reduzindo recursos para atributos contínuos, que pode atender a actividade física e / ou o status de uma pessoa (caminhar, saltar, correr, etc.) de forma mais eficientemente por giroscópio ou acelerómetro de forma independente.

Este artigo serve como uma base para o trabalho futuro com a abordagem de outros métodos. Pode-se citar como exemplos o algoritmo de ganho de informação melhoria através da introdução da dependência de atributos. Então, você começa a recolher dados não só de uma pessoa, mas várias ao mesmo tempo e processamento de informações em tempo real.

## Bibliografía

- 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.