

Modelos de clasificación en marcha patológica usando árboles de regresión logística

Ana Aguilera y Alberto Subero

*Centro de Análisis, Modelado y Tratamiento de Datos, CAMYTD,
Facultad Experimental de Ciencia y Tecnología, Universidad de Carabobo, Venezuela*
aaguilef@uc.edu.ve

Resumen

La Hemiplejía Espástica (HE) es un tipo de parálisis cerebral, en donde los miembros superior e inferior del mismo lado están comprometidos. El Dr. Gage ha sugerido que esta patología pudiera clasificarse en al menos cuatro grupos considerando el patrón cinemático en 3 planos (sagital, transversal y coronal). Tradicionalmente, esta clasificación de pacientes patológicos es realizada por médicos especialistas basándose en el estudio físico del paciente y en el análisis clínico de la marcha, esto último, mediante exámenes complementarios reportados sobre registros cinéticos, cinemáticas y electromiografía. La aplicación de técnicas automáticas de clasificación mediante métodos computarizados constituye un soporte en esta tarea, no como reemplazo al especialista sino como una herramienta de apoyo al diagnóstico. En este trabajo, se analizan diferentes algoritmos de clasificación automática mediante aprendizaje supervisado de los registros de marcha en una población de 255 pacientes recopilados durante seis años. Los árboles de regresión logística combinados con meta-clasificadores han demostrado una buena efectividad en la clasificación. Esta efectividad se evalúa de manera cuantitativa con medidas de especificidad, cantidad de aciertos, falsos positivos y negativos, entre otras medidas. Una validación con médicos especialistas permite comparar los resultados obtenidos por los modelos automáticos y los diagnósticos reales de experto.

Palabras clave: árboles de clasificación logística, marcha patológica, hemiplejía espástica, minería de datos.

Classification Model for Pathological Gait Using Logistic Model Tree

Abstract

Spastic Hemiplegia (SH) is a type of cerebral palsy affecting the upper and lower limbs on the same side of the body. Dr. Gage has suggested that this pathology could be classified into at least four groups considering the kinematic pattern on three planes (sagittal, transverse and coronal). Traditionally, this classification of pathological patients is done by specialized physicians based on physical study and the clinical gait analysis of patients, which uses complementary tests reported on kinetic, kinematic and electromyographical records. The application of automatic classification techniques using computerized methods is a support for this task, a diagnostic support tool, not a replacement for the specialist. This paper analyzes different automatic classification algorithms using supervised learning on gait records from a population of 255 patients collected over six years. Logistical regression trees combined with meta-classifiers have demonstrated good efficacy in the classification task. This effectiveness is evaluated quantitatively by measuring specificity, number of correct classifications, false positives and negatives, among other measurements. Validation with specialized physicians makes it possible to compare results obtained by the automatic models with real expert diagnoses.

Keywords: logistical classification trees, pathological gait, spastic hemiplegia, data mining.

1. Introducción

Partiendo de lo establecido por Landwehr (2003) los árboles de regresión logística (LMT, logistic model tree, en inglés) son un método de aprendizaje supervisado que implementa las técnicas combinadas de regresión logística lineal y árboles de inducción. El método de regresión logística crea un modelo lineal con los datos, con un proceso de ajuste del mismo muy estable. En este modelo se registra una baja varianza con un potencial alto sesgo. Mientras que los árboles de inducción exhiben un sesgo bajo, pero a menudo presentan una alta varianza. Por su parte Loh (2006) afirma que éste último permite capturar los patrones no lineales en los datos, pero es menos estable y con tendencia a sobre ajuste. Es una idea natural combinar estos dos métodos de aprendizaje, basándose en modelos de regresión simple y añadiéndole una más compleja estructura de árbol. En las predicciones de una variable numérica utilizando esta unión de modelos, que es un árbol de decisión con modelos de regresión lineal en las hojas, se han obtenido buenos resultados.

Perry (1992), Sutherland (2002), Vilorio (2003) y Villa (2008) coinciden que el análisis clínico de la marcha es un

método complementario de apoyo al diagnóstico ortopédico, que toma en consideración la biomecánica de la marcha humana mediante varias evaluaciones clínicas no invasivas, permitiendo cuantificar el compromiso neurológico reflejado mediante un patrón de marcha patológica. Cabe destacar que el Dr. Gage es uno de los especialistas que ha tratado de incorporar el análisis clínico de la marcha como instrumento de diagnóstico. Esto se ha llevado a cabo con la idea de comprender los mecanismos involucrados en el estudio de patrones que pudieran identificar los procedimientos quirúrgicos ó conservadores a seguir, y aumentar así la calidad de vida de los pacientes con marcha patológica (Gage, 1991). El análisis comprende: el examen físico articular del paciente y el estudio del video bidimensional, de los registros de cinética, cinemática y electromiografía dinámica.

El análisis observacional, se realiza grabando la marcha del paciente en los planos sagital y frontal, logrando tener mayor detalle de la marcha al reproducirla en cámara lenta. Para Vivas (2005) la cinemática se ocupa del estudio del movimiento sin atender a quien lo produce, se logra colocando marcadores reflectivos en las articulaciones de los miembros inferiores, sobre los cuales rebota un haz de luz

infrarroja que disparan de 5 a 6 cámaras colocadas alrededor del paciente. Se obtienen gráficas de la ubicación espacial y del movimiento de las mismas, en cada fase o subfase de la marcha esto según Rupcich (1991). Por su parte Beseler (2007) plantea que la cinética estudia quién o qué produce el movimiento. Los músculos son los generadores internos del movimiento que interactúan durante la marcha con fuerzas externas como la inercia, gravedad, fuerzas de torque, de reacción del suelo, entre otros. La cinética consta de dos parámetros: potencia y momentos articulares, facilitando más información sobre la causa del movimiento.

Por su parte, la electromiografía dinámica, permite registrar la actividad de hasta diez grupos musculares simultáneamente por medio de electrodos de superficie. Este parámetro, conjuntamente con la cinética, permite separar en forma objetiva las desviaciones primarias (producidas directamente por la lesión neurológica) de aquellas compensatorias (mecanismos usados por el individuo para superar las primeras).

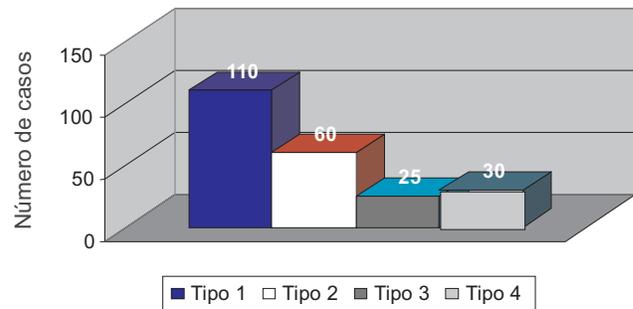
Para lograr diagnósticos y tratamientos más precisos de patologías, los especialistas analizan todos estos registros de marcha para identificar los signos que orientan sus decisiones (Aguilera, 2010).

El objetivo del presente trabajo es la aplicación de la técnica de árboles de regresión logística para sugerir modelos de aproximación diagnóstica en pacientes patológicos con hemiplejia espástica (HE), siguiendo la clasificación propuesta por el Dr. Gage (1991) y (2004) y Winters (1989). Este trabajo está basado en el estudio de los registros de cinética y cinemática.

2. Metodología

2.1. Datos de estudio

Los datos correspondientes a los registros usados en este trabajo fueron recopilados del Laboratorio de Marcha del Hospital Ortopédico Infantil (FHOI, 2006) ubicado en la ciudad de Caracas, Venezuela. Dichos datos provienen del sistema de adquisición VICON370, recopilados entre 1998 y 2004, y corresponden a 278 pacientes con HE. De esta población se consideraron 225 registros cinemáticos almacenados en formato GCD (Gait Cycle Data) que contaban con un diagnóstico tipificado de la patología, por parte del personal médico del laboratorio según el patrón sugerido por el Dr. Gage (1991). La distribución de los casos se muestra en la Figura 1. De la muestra total, se emplearon para la búsqueda y generación de los modelos de clasificación un 75 % de los datos para entrenamiento y se dejó un 25% para pruebas.



Fuente: Base de Datos del Laboratorio de Marcha del FHOI

Figura 1. Distribución de Casos de Hemiplejia Espástica Empleados.

2.2. Métodos

Ya que se disponen de datos suficientes para experimentación, se emplean técnicas de minerías de datos para explorar los datos con miras a encontrar patrones y modelos entre ellos. Se emplearon diferentes técnicas para la exploración de algoritmos de clasificación (Hernández, 2004). Se utilizó la herramienta de aprendizaje automático WEKA (2010) en su versión 3.6, con un aprendizaje supervisado o predictivo, dado que para la tarea de clasificación ya se conoce el tipo de HE al que pertenecen. Esto último quiere decir, que existe una clasificación previa establecida por los especialistas sobre los casos considerados, de modo que es posible separar los datos para entrenar/generar el modelo y otra parte para pruebas. WEKA tiene una gran cantidad de algoritmos implementados para técnicas de aprendizaje para clasificación y regresión (árboles de decisión, reglas, clasificadores bayesianos, máquinas de soporte vectorial, regresión lineal y logística, perceptron multicapas y métodos de vecinos más cercanos).

Tal como lo plantea Villena (2010) la herramienta implementa varios métodos de búsqueda a través del espacio de subconjunto de atributos, también como medidas de evaluación para atributos y subconjuntos de atributos. Las medidas de evaluación incluyen criterios basados en correlación y entropía al igual que rendimiento del esquema de aprendizaje seleccionado para un conjunto particular de subconjunto de atributos. La ventaja es que se reducen los atributos iniciales a aquellos realmente discriminantes entre clases, lo que facilita la implementación y consulta de los modelos generados por aprendizaje.

3. Resultados y discusión

3.1. Estudio de registros cinéticos

De estos registros se tomaron en consideración de ambas piernas las medidas espacios temporales: Apoyo

Simple, Longitud de Zancada, Tiempo de Paso, Cadencia, Velocidad de la Marcha y Doble Soporte. Los 51 puntos, tomados en cada 2% del ciclo de marcha, de las gráficas de cada pierna: Flexión/Extensión de Cadera, Flexión/Extensión de Rodilla, Dorsi-Flexión Plantar de Tobillo, Momento Flexión/Extensión de Cadera, Momento Flexión/Extensión de Rodilla, Momento Dorsi-Flexión Plantar de Tobillo, Potencia de la Articulación de Cadera, Potencia de la Articulación de Rodilla, Potencia de la Articulación de Tobillo, teniendo así 931 variables significativas en el estudio. Al aplicar los algoritmos se seleccionaron los que obtuvieron mayor cantidad de instancias clasificadas correctamente; estos algoritmos fueron: LMT, J48, FT, Random Forest, Random Tree, BFTree, SMO, RBNetwork y MultiLayer Perceptron (WEKA, 2010). El modo de evaluación aplicado a estos clasificadores, es decir, el resultado de aplicar el algoritmo de clasificación comparando la clase predicha con la clase real de las instancias, se realizó

por medio de la validación cruzada. Se realizaron 3 tipos de pruebas con los 9 clasificadores: con todos los atributos, empleando tres tipos de reducción de atributos y la incorporación de metaclasificadores. El mejor modelo de clasificación de la HE de acuerdo al número de aciertos, sensibilidad, especificidad, verosimilitud positiva y negativa, verdaderos positivos, verdaderos negativos, falsos positivos y falsos negativos, fue el árbol LMT con reducción a 25 atributos y ensamblado con el metaclasificador AdaBoosting (Tabla 1).

3.2. Estudio de registros de cinemática

Los parámetros considerados en el estudio corresponden a datos provenientes de los dos miembros inferiores, incluyendo miembro afectado o no, conocidos como miembro ipsilateral y contralateral. Estos parámetros son los siguientes: Apoyo simple, Longitud de paso, Velocidad de marcha, Longitud de ciclo o zancada, Cadencia, Doble

Tabla 1. Resultados obtenidos en WEKA usando registros de cinética.

Clases	Mejores clasificadores									
	RF (913)	SMO (913)	RF (57)	MLP (57)	MLP (27)	MLP (25)	MLP (m57)	MP (m27)	LMT (m25)	
Tipo 1	S	0,907	0,918	0,907	0,887	0,856	0,856	0,887	0,887	0,928
	E	0,822	0,921	0,871	0,921	0,871	0,901	0,921	0,941	0,891
	LR+	5,095	11,62	7,031	11,227	6,635	8,646	11,23	15,03	8,513
	LR-	0,113	0,086	0,106	0,122	0,165	0,159	0,123	0,120	0,081
	ROC	0,926	0,922	0,943	0,954	0,905	0,907	0,92	0,948	0,954
Tipo 2	S	0,818	0,909	0,927	0,909	0,855	0,909	0,891	0,891	0,855
	E	0,909	0,958	0,909	0,937	0,923	0,93	0,056	0,93	0,965
	LR+	8,989	21,64	10,186	14,428	11,103	12,985	0,943	12,73	24,43
	LR-	0,20	0,046	0,0803	0,0971	0,157	0,097	1,946	0,117	0,151
	ROC	0,935	0,959	0,965	0,977	0,967	0,961	0,924	0,959	0,961
Tipo 3	S	0,762	1	0,857	0,952	0,905	0,905	0,952	1	0,905
	E	0,989	0,972	1	0,983	0,977	0,983	0,983	1	0,989
	LR+	69,272	35,71	0	56	39,347	53,235	56	0	82,27
	LR-	0,240	0,028	0,143	0,0488	0,097	0,096	0,048	0	0,096
	ROC	0,992	0,986	0,989	0,999	0,981	0,977	0,989	1	0,996
Tipo 4	S	0,64	0,68	0,6	0,68	0,56	0,68	0,76	0,88	0,84
	E	1	0,988	1	0,971	0,96	0,965	0,971	0,977	0,923
	LR+	0	56,67	0	23,448	14	19,428	26,21	38,26	10,91
	LR-	0,36	0,017	0,4	0,329	0,458	0,331	0,247	0,122	0,173
	ROC	0,974	0,87	0,881	0,88	0,817	0,863	0,892	0,974	0,947

RF (913) = algoritmo (alg.) RandomForest con todos los atributos, SMO (193)= alg. SMO con todos los atributos, RF (57) = alg. RandomForest con 57 atributos, MLP (#) = alg. MultiLayerPerceptrón con # atributos, MLP (m#) = alg. Adaboosting+MultiLayerPerceptrón con # atributos, LMT (m25)= Adaboosting+LMT con 25 atributos.

Fuente: autores.

Soporte, Oblicuidad de la pelvis, Aducción-abducción de la cadera, Valgus/varus de rodilla, Ante/Retro pulsión pélvica, Flexión/Extensión de cadera, Flexión/Extensión de rodilla, Dorsiflexión/Plantiflexión de tobillo, Rotación pélvica, Rotación de cadera y Progresión del pie. El número de instancias estudiadas fueron 198; el número de atributos provenientes de los parámetros de marchas con sus respectivos valores fue un total de 1033 atributos.

Usando un método automatizado de búsqueda de subconjuntos de atributos más apropiados para “explicar” un atributo objetivo, en un sentido de clasificación supervisada, se estudió qué subconjuntos de atributos son los que mejor pueden clasificar la clase de la instancia. Los métodos estudiados están basados en algoritmos de Evaluador del Atributo (Attribute Evaluator –filtres and wrappers–) y Método de Búsqueda (Search method). Los mejores resultados se obtuvieron empleando los siguientes algoritmos: CfsSubsetEval y FilteredAttributeEval que pertenecen al Evaluador del Atributo; BestFirst, LinearForwardSelection, GreedyStepwise, Ranker, que corresponden al Método de Búsqueda. Se emplearon además una serie de algoritmos simples y combinados para la elaboración y evaluación del modelo considerando las variables producto del proceso de selección: Oblicuidad de la pelvis, Aducción-abducción de la cadera, Valgus/varus de rodilla,

Ante/Reto pulsión pélvica, Flexión/Extensión de cadera, Flexión/Extensión de rodilla, Dorsiflexión/Plantiflexión de tobillo, Rotación pélvica, Rotación de cadera, Progresión del pie, Apoyo simple (s), Longitud de paso (m), Velocidad de marcha (m/s), Longitud de ciclo o zancada (m), Cadencia (pasos/min) y Doble soporte (s). Los mejores resultados se obtuvieron usando los siguientes algoritmos: LMT, RBFNetwork, RandomForest, J48, SimpleCart y el metaclassificador AdaBoost.M1), siendo el J48 ensamblado con el AdaBoost.M1 los que arrojaron mejores resultados (Tabla 2).

3.3. Validación

Según Bermejo (2001) los métodos de aprendizaje permiten construir modelos o hipótesis a partir de un conjunto de datos, o evidencia. En la mayoría de los casos, es necesario evaluar la calidad de las hipótesis de la manera más exacta posible, para poder llevar a cabo la aplicación en el mundo real de la minería de datos. Para ello, existen varios métodos que sirven para evaluar la calidad de un modelo a partir de la evidencia. La validez o exactitud de un modelo es el grado en el que éste clasifica lo que tiene que clasificar. Es la ausencia de errores sistemáticos.

La fiabilidad es el grado de estabilidad conseguido cuando se repite la prueba en condiciones similares. Las

Tabla 2. Mejores clasificadores para cinemática.

Clase		Mejores Clasificadores											
		LMT(TI)	RBF(67)	RF(67)	RF(69)	LMT(30)	RF(30)	RF(78)	RF(M-67)	J48(M-69)	RF(M-30)	J48(M-78)	SC(M-78)
1	S	0,896	0,907	0,896	0,938	0,855	0,938	0,907	0,938	0,938	0,917	0,958	0,907
	E	0,851	0,891	0,851	0,981	0,871	0,981	0,881	0,851	0,841	0,9	0,9	0,881
	LR+	6,039	8,32	6,039	8,613	6,647	8,613	7,635	6,316	5,922	9,267	9,683	7,635
	LR-	0,121	0,104	0,121	0,069	0,165	0,069	0,115	0,072	0,073	0,091	0,045	0,105
	ROC	0,928	0,918	0,959	0,973	0,922	0,947	0,948	0,955	0,961	0,946	0,951	0,962
2	S	0,745	0,818	0,8	0,89	0,8	0,89	0,818	0,8	0,818	0,909	0,8	0,89
	E	0,951	0,909	0,902	0,923	0,916	0,923	0,937	0,944	0,951	0,93	0,951	0,933
	LR+	15,228	9	8,171	11,581	9,533	11,581	10	14,3	16,714	13	16,571	13,3636
	LR-	0,267	0,2	0,22	0,118	0,218	0,118	0,194	0,211	0,191	0,097	0,21	0,116
	ROC	0,941	0,889	0,933	0,979	0,868	0,951	0,938	0,95	0,956	0,951	0,925	0,969
3	S	1	0,761	0,761	0,571	0,904	0,857	0,904	0,952	0,857	0,809	0,904	0,857
	E	0,988	0,988	0,938	1	0,994	1	0,988	1	0,994	0,994	0,988	1
	LR+	88,5	67,428	44,952	---	160,142	---	80,071	---	151,714	143,285	79,619	---
	LR-	0	0,24	0,242	0,142	0,095	0,142	0,096	0,047	0,147	0,191	0,096	0,142
	ROC	0,99	0,949	0,994	0,98	0,974	0,992	0,991	1	0,988	0,979	0,996	0,997
4	S	0,8	0,72	0,72	0,72	0,84	0,72	0,84	0,76	0,72	0,76	0,84	0,8
	E	0,971	0,971	0,994	1	0,971	1	0,988	0,994	0,988	0,988	0,988	0,98
	LR+	27,68	24,912	124,56	---	29,064	---	72,66	131,48	60,84	65,74	72,66	69,2
	LR-	0,205	0,288	0,281	0,142	0,164	0,28	0,161	0,241	0,283	0,242	0,161	0,202
	ROC	0,935	0,854	0,961	0,959	0,961	0,984	0,98	0,953	0,961	0,962	0,988	0,951

Fuente: autores.

medidas comúnmente usadas son el índice de concordancia global, el índice de concordancia Kappa, el coeficiente de variación y el coeficiente de correlación interclases. Para la validez de un modelo se mide la sensibilidad y la especificidad. Las curvas ROC (Receiver Operating Characteristic) son una herramienta visual muy útil a la hora de comparar dos modelos de clasificación. En ellas se muestra la comparación entre la tasa de verdaderos positivos o sensibilidad y la tasa de falsos positivos de un modelo dado. Un modelo con exactitud perfecta tendrá un área de 1.0. Tan importantes son la fiabilidad y la validez, que la ausencia de ellas conlleva a un error que se trata de minimizar (Bermejo, 2001).

Es posible tratar dos tipos de validación para la evaluar la efectividad de los modelos obtenidos, una cualitativa mediante validación y verificación directa de los expertos, y otra cuantitativa mediante técnicas de validación existentes como medición de fiabilidad, índice de kappa, curvas ROC y otras medidas de validez como sensibilidad, especificidad y razón de verosimilitud. Se distingue la validación cualitativa como aquella apoyada por la observación

directa de expertos y que no emplea métodos estadísticos u otros numéricos.

3.3.1. Cualitativa

En la sección 2a se indicó que un 25% de la población bajo estudio se utilizó para probar la efectividad del modelo. Esto contrasta la clasificación emitida automáticamente por el modelo y la clasificación existente del experto. Adicionalmente se realizó una selección de un grupo de pacientes con patología de HE tratados en el laboratorio de marcha de la FHOI. Se presentaron a los especialistas los videos bidimensionales de cada paciente en dos planos (Sagital y Frontal) y las respectivas gráficas de cinemática y cinética. Luego, los resultados emitidos por cada clasificador automático fueron comparados con los diagnósticos emitidos por los médicos de manera habitual. Fueron consultados 3 médicos expertos en el área, se utilizaron 5 pacientes cada uno con 3 ó 4 triales (o recorrido de marcha), para un total entre 15 y 20 estudios analizados. La Tabla 3 muestra de manera tabular un resumen de los resultados obtenidos. En esta tabla se presentan los triales de los pa-

Tabla 3. Resultados de clasificación automática y de expertos de los registros de cinética y cinemática.

Triales Evaluados	Diagnóstico Dr. Prato y Colaboradores	Aprox. Cinemática (%)		Aprox. Cinética (%)	
		Tipo HE (%)	Clasif	Tipo HE (%)	Clasif
50404	Tipo 4	Tipo 1: 13,3	Tipo 4	Tipo 1: 62,5	Tipo 4
50405		Tipo 2: 11,4		Tipo 2: 53,75	
50406		Tipo 3: 0		Tipo 3: 0,03	
50408		Tipo 4: 75,3		Tipo 4: 76,25	
62008	Tipo 4	Tipo 1: 14,5	Tipo 4	Tipo 1: 0	Tipo 2
62010		Tipo 2: 17,1		Tipo 2: 70	
62015		Tipo 3: 24,16		Tipo 3: 53	
		Tipo 4: 44,24		Tipo 4: 61,67	
50707	Tipo 1	Tipo 1: 96,81	Tipo 1	Tipo 1: 91,25	Tipo 1
50713		Tipo 2: 31,19		Tipo 2: 52,42	
50716		Tipo 3: 0		Tipo 3: 4,53	
50720		Tipo 4: 0		Tipo 4: 51,2,3	
51403	Tipo 1	Tipo 1: 93,62	Tipo 1	Tipo 1: 85	Tipo 1
51407		Tipo 2: 6,38		Tipo 2: 60	
51408		Tipo 3: 0		Tipo 3: 0	
		Tipo 4: 0		Tipo 4: 61,25	
36517	Tipo 1	Tipo 1: 90,87	Tipo 1	Tipo 1: 90	Tipo 1
36519		Tipo 2: 4,36		Tipo 2: 52,41	
36520		Tipo 3: 0		Tipo 3: 2,34	
36523		Tipo 4: 4,77		Tipo 4: 57,5	

Fuente: autores.

cientes que fueron seleccionados junto a los resultados emitidos por cada clasificador y el diagnóstico consolidado de los médicos.

Es de resaltar que de los cinco pacientes evaluados, todos los resultados emitidos por el clasificador de cinemática coincidieron en un 100%, mientras que el modelo cinético solo obtuvo un resultado en discordancia con el diagnóstico de expertos. Sin embargo, el segundo porcentaje más alto arrojado se inclina al Tipo 4.

3.3.2. Validación cuantitativa

La validación cuantitativa de los modelos resultantes por cada registro o señal, para cada clase (Tipo 1 al 4) de HE, incluye las mediciones arrojados por el Índice de Kappa (IK), las Instancias Clasificadas Correctamente (ICC), Verosimilitud Positiva (LR+), Verosimilitud Negativa (LR-), Sensibilidad (S), Especificidad (E) y ROC = Área bajo la Curva.

Cinética

En las pruebas anteriores se pudo observar que en las comparaciones y validaciones de los mejores modelos el que dio resultados más satisfactorios fue el LMT ensamblado con el metaclasificador Adaboosting con una reducción a 25 atributos, el cual tiene uno de los mayores porcentajes de satisfacción (89,3939%) y un error medio de 0,0576; su índice de concordancia en los datos es muy bueno con 0,837 (Tabla 4).

Cinemática

De acuerdo a las mediciones descritas anteriormente, el algoritmo que mejor dio resultados en cinemática, en

porciones equilibradas para cada una de las clases, fue el J48 ensamblado con el AdaBoost.M1 con reducción a 78 atributos.

Los criterios utilizados toman en cuenta la fiabilidad del modelo: Instancias Clasificadas Correctamente 177 de 198 lo que representa en porción un 89,3939%, Índice de Kappa = 0,8362. El proceso de selección redujo 1033 atributos a 78. Estos atributos resultantes representan los 78 primeros con mayor rango en aportar información relevante y clave para los algoritmos FilteredAttributeEval y Ranker. El J48 utiliza la técnica de ponderación que sustituye de una mejor manera aquellas ramas del árbol con poca capacidad predictiva para reducir el ruido de los datos de entrenamiento. Por su parte, el AdaBoostM1 al derivar del algoritmo Boosting le asigna un peso (número positivo) específico a cada árbol o modelo que determina el grado de significancia que tiene el modelo. A mayor peso por instancias, dicho algoritmo toma más interés por clasificarlas correctamente (Tabla 5).

4. Consideraciones finales

Los algoritmos de clasificación en HE propuesto para cinética y cinemática logran discriminar entre los cuatro tipos de la patológica, del cual se obtuvo un 89,39% de instancias clasificadas correctamente, con medidas de validez por encima del 84%. La reducción de atributos de 913 a 25 y de 1033 a 78, en el caso de la cinética y cinemática respectivamente, hace mucho más sencillo y preciso el modelo, pues solo son tomadas en consideración aquellas variables significativas que mejor representan cada clase.

Tabla 4. Medidas de efectividad del LMT para cinética.

	Class	Type 1	Type 2	Type 3	Type 4
Classifier LMT-M-25	S	0,928	0,855	0,905	0,84
	E	0,891	0,965	0,989	0,923
	LR+	8,513	24,428	82,272	10,909
	LR-	0,0808	0,1502	0,096	0,173
	ROC	0,954	0,961	0,996	0,947

=== Stratified cross-validation ===

=== Summary ===

Correctly Classified Instances	177	89,3939 %
Incorrectly Classified Instances	21	10,6061 %
Kappa statistic	0,837	
Mean absolute error	0,0576	
Root mean squared error	0,2305	
Relative absolute error	17,5005 %	
Root relative squared error	56,9021 %	
Total Number of Instances	198	

Fuente: autores.

Tabla 5. Medidas de efectividad del J48 para cinemática.

	Class	Type 1	Type 2	Type 3	Type 4
Classifier J48-M-78	S	0,958	0,8	0,904	0,84
	E	0,9	0,951	0,988	0,988
	LR+	9,683	16,571	76,619	72,66
	LR-	0,045	0,21	0,096	0,161
	ROC	0,951	0,925	0,996	0,988

=== Stratified cross-validation ===

=== Summary ===

Correctly Classified Instances	177	89,3939 %
Incorrectly Classified Instances	21	10,6061 %
Kappa statistic	0,8362	
Mean absolute error	0,0552	
Root mean squared error	0,221	
Relative absolute error	16,7802 %	
Root relative squared error	54,5803 %	
Total Number of Instances	198	

Fuente: autores.

Es muy importante resaltar que los especialistas al realizar el análisis clínico de la marcha, específicamente en pacientes HE, basan su clasificación tipológica principalmente mediante los registros de cinemática (según la clasificación propuesta por Gage, 1991). Sin embargo los registros de cinética, aportan información importante para esta clasificación (Aguilera, 2011). De este modo, el estudio de modelos automáticos que sugieran una clasificación tipológica de la HE, tomando en consideración los otros registros de estudio como la cinética y la EMG como elemento determinista, señalarían información adicional sobre las variables y señales involucradas aportando información discriminante entre los tipos de HE.

El aporte principal de este trabajo es la utilización de técnicas combinadas de minería de datos (metaclasificadores, árboles de regresión logística y J48) sobre registros cinéticos y cinemáticas con validaciones cuantitativas importantes en base a criterios como sensibilidad (>84%) y especificidad (>89,1%).

Agradecimientos

Este trabajo ha sido subvencionado por el Consejo de Desarrollo Científico y Humanístico de la Universidad de Carabobo (CDCH-UC) según oficio Nro. CDCH-AM-0465-10, del 12-11-2010. Los autores agradecen al personal del Laboratorio de Marcha del FHOI, especialmente a la fisioterapeuta Danirida Urbano y al Dr. Carlos Prato jefe del Laboratorio de marcha por su colaboración.

Referencias

- AGUILERA, Ana; CALA, Luz., VILORIA, Ninoska (2011). Aplicación de Apoyo al Diagnóstico en Marcha Patológica Mediante Análisis Cinético. *Revista FARAUTE*. Latindex Venezuela, Volumen 6, número 1.
- BERMEJO, Begoña (2001). Epidemiología Clínica aplicada a la toma de decisiones en medicina. WWW:http://www.cfnava.es/salud/docencia.investigacion/textos/Monograf_1/Epidemiologia_clinica.pdf (Consulta: 31/10/2011).
- BESELER, María (2007). Estudio de los parámetros cinéticos de la marcha del paciente hemipléjico mediante plataformas dinámicas. Tesis Doctoral Universidad de Valencia, Facultad de Medicina, España.
- DIONISIO, Valdeci; ALMEIDA, Gil; DUARTE, Marcos; HIRATA, Rogério (2006). Kinematic, kinetic and EMG patterns during downward squatting. *Journal of Electromyography and Kinesiology*.
- FHOI, Fundación Hospital Ortopédico Infantil (2006). Disponible: http://www.ortopedicoinfantil.org/hospital/servicios/auxiliares_diagnostico/laboratorio_marcha.htm/. (Consulta: 13/6/2010)
- GAGE, James (1991). *Gait Analysis in Cerebral Palsy*. Mac Keith Press, 1st Edition.
- GAGE, James (2004). A qualitative description of normal gait. In: Gage JR, ed. *The Treatment of Gait Problems in Cerebral Palsy*. London: Mac Keith Press.
- HERNÁNDEZ, José; RAMÍREZ, José; FERRI, César (2004). *Introducción a la minería de datos*. Madrid: Pearson Educación.
- LANDWEHR, Niels; HALL, Mark; FRANK, Eibe (2003). Logistic Model Trees. Disponible: <http://www.cs.waikato.ac.nz/~ml/publications/2003/landwehr-etal.pdf>. (Consulta: 13/6/2010).

- LOH, Wei-Yin (2006). Logistic Regression Tree Analysis. Handbook of Engineering Statistics, H. Pham, ed., 537-549, Springer. Disponible: <http://www.stat.wisc.edu/~loh/tree-progs/lotus/springer.pdf>. (Consulta: 13/6/2010).
- PERRY, Jacquelin (1992). **Gait Analysis, Normal and Pathological Function**. McGraw-Hill, 2nd Edition, USA.
- RUPCICH, Marcel (1991). El Laboratorio de Marcha y su Utilidad Clínica". La Fundacion Hospital Ortopédico Infantil. Caracas. Venezuela. Disponible en: [http://www.dynabizvenezuela.com/images/dynabiz/ID3749/siteinfo/El Laboratorio de Marcha.pdf](http://www.dynabizvenezuela.com/images/dynabiz/ID3749/siteinfo/El_Laboratorio_de_Marcha.pdf). (Consulta 19/6/2010).
- SUTHERLAND, David (2002). The evolution of clinical gait analysis Part II Kinematics. Elseiver - Gait and Posture, 16.
- VICON370 Documentación del Sistema. Disponible en: http://www.vicon.com/applications/gait_analysis.html. Consultado 31/10/11.
- VILLENA, Julio (2010). "Técnicas de Análisis de Datos en WEK". Disponible en: www.it.uc3m.es/jvillena/irc/md/tutorialWeka.pdf. Consultado 19/6/2010.
- VILLA, Adriana; GUTIÉRREZ, Eduardo; PÉREZ, Juan (2008). Consideraciones para el análisis de la marcha humana. Técnicas de videogrametría, electromiografía y dinamometría. **Revista Ingeniería Biomédica**. volumen 2, número 3, enero-junio 2008.
- VILORIA, Ninoska (2003). Evaluación Electromiográfica de la clasificación cinemática de Hemipléjicos Espásticos con marcha patológica. Maestría en Ingeniería Biomédica Tesis de Grado, Universidad Simón Bolívar, Caracas, Venezuela.
- VIVAS, María; BAYDAL, José; PEYDRO, María; GARRIDO, José (2005). Contribución del análisis cinético de la marcha a la valoración de los trastornos del equilibrio. **Revista de Biomecánica** del Instituto de Biomecánica de Valencia, (44) Julio 2005.
- WEKA (2010). Data Mining Software in Java. The University of Waikato. Disponible: <http://www.cs.waikato.ac.nz/ml/weka/>. (Consulta: 13/6/2010).
- WINTERS, Tomas; GAGE, James; HICKS, Ramona (1987). Gait Patterns in spastic hemiplegia in children and young adults. **Journal of Bone Joint Surgery**.