

Análisis comparativo de algoritmos de aprendizaje para predecir la evolución de pacientes con Daño Cerebral Adquirido

A. Marcano-Cedeño^{1,2}, P. Chausa^{1,2}, C. Cáceres^{1,2}, A. García³, R. López³, J.M. Tormos³, E.J. Gómez^{1,2}

¹ Grupo de Bioingeniería y Telemedicina,ETSIT- Universidad Politécnica de Madrid, España
{amarcano,pchausa,ccaceres,egomez}@gbt.tfo.upm.es

² Centro de Investigación Biomédica en Red en Bioingeniería, Biomateriales y Nanomedicina (CIBER-BBN), Madrid, España

³ Servicio de Investigación, InstitutGuttmann, Badalona, España. {agaciar,rlopez,jmtormos}@guttmann.com

Resumen

Este trabajo presenta un análisis comparativo entre tres algoritmos de aprendizaje diferentes basados en Árboles de Decisión (C4.5) y Redes Neuronales Artificiales (Perceptrón Multicapa MLP y Red Neuronal de Regresión General GRNN) que han sido implementados con el objetivo de predecir los resultados de la rehabilitación cognitiva de personas con daño cerebral adquirido. En el análisis se han incluido datos demográficos del paciente, el perfil de afectación y los resultados provenientes de las tareas de rehabilitación ejecutadas por los pacientes. Los modelos han sido evaluados utilizando la base de datos del InstitutGuttmann. El rendimiento de los algoritmos se midió a través del análisis de la especificidad, sensibilidad y exactitud en la precisión y el análisis de la matriz de confusión. Los resultados muestran que la implementación del C4.5 alcanzó una especificidad, sensibilidad y exactitud en la precisión del 98.43%, 83.77% y 89.42% respectivamente. El rendimiento del C4.5 fue significativamente superior al obtenido por el Perceptrón Multicapa y la Red de Regresión General.

1. Introducción

El Daño Cerebral Adquirido (DCA) altera la vida diaria del paciente y de las personas de su alrededor a consecuencia de las alteraciones físicas, sensoriales, emocionales y cognitivas que produce. Estas últimas generan una mayor dependencia, ya que dificultan la actividad, interacción y participación en el entorno familiar, laboral o social. La etiología puede ser traumática (Traumatismo Cranoencefálico, TCE) o no traumática (Accidente Cerebrovascular, ACV). En España, cada año, se producen 2700 TCEs (entre graves y moderados) que desembocarán en situaciones de discapacidad cognitiva, tributarios de recibir tratamiento rehabilitador. A esto hay que sumar los casos de Ictus, que afectan a un número todavía mayor de personas.

La rehabilitación neuropsicológica es una disciplina relativamente joven y su cuerpo de conocimiento es fundamentalmente empírico e implícito. El número de estudios en los últimos años ha aumentado de manera exponencial, pero los resultados son poco concluyentes y las revisiones realizadas no permiten establecer una práctica de la rehabilitación basada en la evidencia. La rehabilitación cognitiva es un proceso costoso y los recursos que se pueden dedicar son limitados. Esto hace

necesario valorar fielmente su eficacia para optimizar los procedimientos. Es necesario poder determinar qué, porqué, a quién y cómo debe utilizarse.

Diferentes metodologías estadísticas y técnicas de minería de datos se han aplicado para predecir la eficacia del proceso de rehabilitación de los pacientes con DCA. En este trabajo se analizan tres algoritmos de aprendizaje: un Árbol de Decisión, un Perceptrón Multicapa y una Red Neuronal de Regresión General. Cada algoritmo ha generado un modelo de predicción de la evolución de las funciones cognitivas de personas con DCA después de haber superado un proceso de rehabilitación específico.

2. Modelos de predicción en DCA

El desarrollo de un modelo para estudiar la correlación entre los factores de predicción y determinar la evolución de un paciente con DCA ha sido objeto de diversos estudios y publicaciones. Los autores utilizan diferentes técnicas estadísticas y de minería de datos tratando de encontrar aquella que más se adapta a sus necesidades y proporciona mejores resultados. Podemos encontrar estudios generales que analizan modelos basados en árboles de decisión (C4.5 y CART), metaclasificadores (AdaBoost), SVM (Support Vector Machine) y redes neuronales artificiales [1]. Algunos incorporan además el análisis de modelos de regresión múltiple [2], regresión logística [3], redes bayesianas y análisis discriminante [4] e incluso la predicción proporcionada por neurocirujanos especialistas [5]. De los resultados obtenidos se destaca la buena predicción obtenida por los modelos de árboles de decisión y regresión logística [4]. En el trabajo realizado por Peter et al. [3], el sistema de árboles de decisión confirmó algunos de los resultados obtenidos con regresión logística y puso de manifiesto la existencia de otras variables significativas. Además, los autores destacan la capacidad de interpretación de los resultados obtenidos por los árboles de decisión frente a los métodos de regresión. Soon-Yeon Ji et al. van más allá en su análisis y proponen un sistema de ayuda a la decisión basado en C4.5 y CART capaz de predecir la supervivencia del paciente y la necesidad de un proceso de rehabilitación. En otros casos fue el modelo de red neuronal el que mejoró el comportamiento de los métodos de regresión y la predicción de los especialistas, en

términos de precisión, sensibilidad y discriminación [5], o mostró un valor de predicción similar al modelo de regresión múltiple siendo ambos superiores al CART [2].

Además del uso y la comparación de diferentes técnicas y algoritmos, cada investigación basa su análisis en una combinación diferente de atributos característicos de los pacientes. En [6], los autores se plantean como objetivo identificar cuáles son las variables más significativas a la hora de predecir la evolución de un paciente con TCE. Para ello utilizan árboles de decisión con particionado recursivo. Analizan un gran número de variables: la puntuación FIM (Medida de Independencia Funcional), las puntuaciones de la escala DRS (Disability Rating Scale), la escala SRS (Supervision Rating Scale) y la escala GOS, una medida de la actividad productiva, datos del examen físico del paciente, duración de la amnesia post-traumática, duración del coma, resultados del TAC, sexo, edad y años de educación. Una vez realizado el estudio se comprobó que la duración de la amnesia post-traumática es el mejor predictor seguido de los datos provenientes del examen físico del paciente. Con objetivo y metodología similares, Rovlias y Kotsou utilizan un modelo CART para predecir la evolución de un paciente con TCE severo y aislado a los 6 meses [7]. En este caso los autores pretenden realizar un sistema de predicción simple que utilice variables que se puedan adquirir rápida y fácilmente en la rutina práctica diaria. Analizaron datos de 345 pacientes y examinaron un total de 16 indicadores. Como resultado, el estudio extrajo que la puntuación de la escala GOS era la variable más significativa. Otras variables que destacaron por su capacidad predictiva fueron la edad, reacción de la pupila, resultados del TAC, hiperglucemia y leucocitosis.

En este trabajo presentamos un estudio comparativo entre métodos que han demostrado proporcionar buenos resultados de predicción: modelos basados en redes neuronales (MLP y GRNN) y árboles de decisión (C4.5). El análisis se ha centrado en la evolución de la función cognitiva memoria en pacientes con DCA. Por ello, como variables de entrada se han seleccionado ítems específicos de la batería neuropsicológica utilizada para valorar a los pacientes además de algunos datos demográficos de los pacientes (edad y nivel de estudios). Un elemento diferenciador de este trabajo es la inclusión como variable de entrada de los resultados obtenidos por los pacientes al ejecutar las tareas de rehabilitación en una plataforma de tele-rehabilitación cognitiva.

3. Materiales y Métodos

3.1. Descripción de los datos

El Institut Guttmann (IG) es un centro de referencia especializado en la rehabilitación integral (tanto psicológica como social) y el tratamiento médico/quirúrgico de las personas afectadas por una lesión medular, un DCA o cualquier otra discapacidad física grave de origen neurológico. Una de las áreas de experiencia del IG es la rehabilitación cognitiva de pacientes con DCA. Al ingresar en el centro los pacientes realizan una batería de pruebas neuropsicológicas para evaluar sus funciones cognitivas (atención, memoria y

función ejecutiva). La batería utilizada por el IG consta de 27 tests que cubren los principales dominios cognitivos. Tras esta evaluación, los pacientes realizan un tratamiento de rehabilitación cognitiva por medio de la plataforma de Tele-rehabilitación Cognitiva PREVIRNEC. El tratamiento tiene una duración promedio de 3 a 5 meses, adecuado al nivel de afectación y evolución de cada paciente. Una vez finalizado el programa de rehabilitación, los pacientes son valorados nuevamente con la misma batería. La diferencia entre las valoraciones pre-tratamiento y post-tratamiento nos indican la evolución del paciente en las distintas funciones cognitivas.

En este estudio se ha utilizado una muestra de 250 pacientes del Institut Guttmann (185 hombres y 65 mujeres) afectados por DCA que se sometieron al tratamiento de rehabilitación con la plataforma PREVIRNEC. La edad media de los pacientes es de 36,56 \pm 6,5 años, con un rango de edad de 18 a 56 años. El análisis se ha centrado en el estudio de la evolución de los pacientes en la función cognitiva memoria. Para ello se han seleccionado los 6 tests de la batería de exploración relacionados con dicha función y su puntuación se ha normalizado en el rango 0-4 según el grado de afectación: normal, leve, moderado, grave o muy grave. De entre todas las tareas de rehabilitación disponibles en la plataforma PREVIRNEC, se han incluido en el estudio los resultados de ejecución de la tarea "Memory".

La base de datos utilizada consta de 1186 registros (número de ejecuciones de la tarea "Memory") de los cuales 719 (60.62%) corresponden a ejecuciones de pacientes que mejoraron y 467 (39.38%) a pacientes que no mejoraron. Para evaluar comparativamente el desempeño de los diferentes métodos presentados en este estudio, todos fueron entrenados con los mismos conjuntos de datos de entrenamiento (60% del total, de los cuales 431 pertenecen a la clase mejora y 280 a la de no mejora) y prueba (40% del total, con 288 muestras de la clase mejora y 187 de no mejora). Cada registro consta de 9 atributos y una etiqueta de clase (ver Tabla 1).

Descripción	Rango de valores
Edad	[17-30 años, 31-56, >56]
Nivel estudios	[Primarios, Secundarios, Universitarios]
MDIGITS	[0, 1, 2, 3, 4]
MLETRES	[0, 1, 2, 3, 4]
SPAN	[0, 1, 2, 3, 4]
RAVL075	[0, 1, 2, 3, 4]
RAVL015	[0, 1, 2, 3, 4]
RAVL015TR	[0, 1, 2, 3, 4]
Resultado	[0-100]
Evolución	[MEJORA, NO MEJORA]

Tabla 1 Descripción de los registros analizados

MDIGITS, MLETRES y SPAN son tres de los tests de la batería de exploración neuropsicológica WAIS-III que se

usa para medir la inteligencia en adultos y adolescentes. RAVL075, RAVL015 y RAVL015TR pertenecen a la batería RAVLT que evalúa el aprendizaje verbal y la memoria.

3.2. Diseño del modelo predictivo

En este estudio se han incluido tres algoritmos de clasificación, C4.5, MLP y GRNN, que han sido implementados con la plataforma de software libre Weka. Para evaluar el rendimiento de los clasificadores se ha utilizado un procedimiento de validación cruzada (10-fold crossvalidation). Los datos fueron divididos en 10 grupos de forma que, en cada iteración, nueve de ellos fueron utilizados para entrenar el modelo y el conjunto restante para probarlo. Los resultados de los 10 experimentos de cada modelo se sumaron para estimar la exactitud en la clasificación de los algoritmos.

Árboles de decisión (DT)

Los árboles de decisión son modelos de clasificación y predicción ampliamente utilizados en minería de datos. Esta técnica separa recursivamente observaciones en ramas para construir un árbol con el propósito de mejorar la exactitud en la predicción. Para realizar la separación de las muestras se utilizan diferentes algoritmos matemáticos como el índice de Gini (utilizado en CART) o un método basado en el aumento de la entropía de la información (utilizado en ID3, C4.5, C5) [8]. En este estudio se ha utilizado el algoritmo J48, una implementación del algoritmo C4.5 de Weka.

Redes Neuronales Artificiales

El MLP es una de las configuraciones de red más populares y ha sido utilizado con éxito en problemas de medicina, de reconocimiento de patrones, predicción, modelado de procesos, etc. El algoritmo de retropropagación (BP) es el algoritmo de aprendizaje comúnmente usado para entrenar a un MLP. Durante la fase de entrenamiento el BP cambia iterativamente los pesos de las neuronas con la finalidad de minimizar el error. El error entre las salidas reales de la red y la deseada se calcula y se utiliza para ajustar los pesos de conexión. El procedimiento de ajuste es derivado de un método de descenso de gradiente y se utiliza para reducir la magnitud del error [9].

La GRNN es un tipo de Red de Función de Base Radial normalizada, donde existe una celda de la capa escondida correspondiente a cada patrón de entrenamiento. Este tipo de redes está basado en la Teoría de Regresión No lineal. Realiza una buena aproximación o mapeo de funciones entrada-salida a partir de los datos de entrenamiento. A medida que el conjunto de entrenamiento crece, el error se aproxima a cero [9]. Es posible considerar a las GRNN como un aproximador universal para funciones suaves. Así, debería ser factible resolver cualquier problema de aproximación de funciones disponiendo de datos suficientes. El inconveniente principal de estas redes, como el de los métodos de kernels en general, es que sufre de "dimensionalidad"; por lo tanto la cantidad de patrones de entrenamiento es un variable limitante para la aplicación de este tipo de predictores [10].

4. Resultados

4.1. Medidas de rendimiento

Los promedios de la exactitud de la clasificación de los tres modelos fueron analizados y comparados con el objetivo de analizar cuál de ellos ofrece una predicción más exacta de la mejora de los pacientes. En general, las medidas de calidad de la clasificación se construyen a partir de una **matriz de confusión** que registra:

- Verdaderos positivos (VP): número de pacientes que mejoraron y fueron clasificados como que mejoraron.
- Falsos positivos (FP): número de pacientes que no mejoraron y fueron clasificados como que mejoraron.
- Falsos negativos (FN): número de pacientes que mejoraron y fueron clasificados como que no mejoraron.
- Verdaderos negativos (VN): número de pacientes que no mejoraron y fueron clasificados como que no mejoraron.

En este trabajo de investigación utilizamos las siguientes medidas de evaluación, definidas a partir de la matriz de confusión:

- Especificidad = $VN / (FP + VN)$
- Sensibilidad = $VP / (VP + FN)$
- Exactitud = $(VP + VN) / (VP + FN + FP + VN)$

4.2. Selección de la estructura de red

Para seleccionar la mejor configuración de cada clasificador, se probaron varias estructuras de red y diferentes parámetros. La Tabla 2 muestra las tres configuraciones que dieron mejor resultado para el caso del árbol de decisión destacándose la que fue seleccionada para el análisis. En las Tablas 3 y 4 se presenta el mismo resultado para el MLP y GRNN.

Algoritmo	Factor de confianza	Nº mínimo de objetos	Exactitud predicción
DT1	0.25	2	90.95 %
DT 2	0.5	2	86.63 %
DT3	0.6	2	80.37 %

Tabla 2. Diferentes parámetros para DT

Algoritmo	HL	LR	Momentum	Épocas	Exactitud predicción
MLP 1	5	0.1	0.1	1000	78.73 %
MLP 2	10	0.9	0.2	2000	76.42 %
MLP 3	15	0.5	0.2	3000	78.73 %

Tabla 3. Estructuras de la red y los parámetros para MLP

La configuración seleccionada para el caso del Perceptrón Multicapa fue la MLP1 debido a su menor coste computacional.

Algoritmo	Propagación	Exactitud predicción
GRNN 1	1	90.95 %

GRNN 2	1.25	86.63 %
GRNN 3	1.3	80.37 %

Tabla 4 Parámetros para las GRNN

4.3. Evaluación del rendimiento

La exactitud en la predicción evalúa la eficacia general del algoritmo mientras que la especificidad y la sensibilidad estiman el rendimiento de las diferentes clases siendo dos indicadores muy utilizados en el dominio clínico [11]. La Tabla 5 muestra los resultados de la clasificación obtenida en la mejor simulación de cada clasificador utilizado en este estudio.

Algoritmo	Predicción		
	Mejora	No Mejora	
DT	Mejora	185	2
	No Mejora	43	245
MLP	Mejora	176	11
	No Mejora	90	198
GRNN	Mejora	110	77
	No Mejora	82	206

Tabla 5. Matriz de confusión de la mejor clasificación alcanzada por cada algoritmo

La Tabla 6 muestra el valor de la media y desviación estándar de las medidas de rendimiento para los tres algoritmos analizados obtenidos a partir del proceso de validación cruzada (10-fold crossvalidation).

	Predicción		
	Especificidad	Sensibilidad	Exactitud
DT	98.43% ± 0.8	83.77% ± 0.9	89.42% ± 0.9
MLP	92.81% ± 1.3	67.25% ± 1.5	77.53% ± 1.2
GRNN	53.32% ± 0.5	75.13% ± 0.4	66.02% ± 0.5

Tabla 6. Matriz de confusión de la mejor clasificación alcanzada por cada algoritmo

5. Discusión y Conclusiones

Los modelos construidos permiten predecir el resultado de la función cognitiva memoria a partir del perfil de afectación inicial del paciente, su edad y nivel de estudios y los resultados provenientes del proceso de rehabilitación. De la comparación realizada entre los tres algoritmos empleados vemos cómo los resultados del árbol de decisión son mejores a los obtenidos por el MLP y el GRNN en términos de especificidad, sensibilidad y exactitud en la predicción (ver Tablas 5 y 6). De igual manera, si comparamos la desviación estándar con la eficiencia de clasificación más estable, el árbol de decisión es de nuevo superior al resto de algoritmos de clasificación.

Esta misma investigación podría extenderse al resto de funciones cognitivas definidas por el Institut Guttmann (atención y función ejecutiva) dando lugar a un modelo de

predicción completo para pacientes tratados mediante la plataforma PREVIRNEC.

Complementariamente, sería interesante analizar los datos de entrada para realizar una asociación entre el perfil de los pacientes y la evolución de los mismos. Esto nos permitiría, además de predecir el resultado de la rehabilitación, conocer las características de los pacientes que mejoran y los programas de rehabilitación que han seguido. Esto generaría un aumento de conocimiento sobre el proceso de rehabilitación y facilitaría la elaboración de hipótesis de los terapeutas sobre el tipo de programa que está relacionado con la mejora de un cierto tipo de pacientes. También sería interesante analizar cada elemento de la batería de exploración neuropsicológica de forma individual con el fin de determinar la capacidad predictiva de cada uno de estos test y la posible relación entre ellos.

Referencias

- [1] Ji, S., Smith, R., Huynh, T. et al.: A Comparative Analysis of Multi-Level Computer-Assisted Decision Making Systems for Traumatic Injuries. *BMC Medical Informatics and Decision Making*, 9 (2009) 2
- [2] Segal, M. E., Goodman, P. H., Goldstein, R. et al.: The Accuracy of Artificial Neural Networks in Predicting Long-Term Outcome After Traumatic Brain Injury. *J. Head Trauma Rehabil.*, 21 (2006)
- [3] Andrews, P. J. D., Sleeman, D. H., Statham, P. F. X. et al.: Predicting Recovery in Patients Suffering from Traumatic Brain Injury by using Admission Variables and Physiological Data: A Comparison between Decision Tree Analysis and Logistic Regression. *J. Neurosurg.*, 97 (2002) 326-336
- [4] Pang, B. C., Kuralmani, V., Joshi, R. et al.: Hybrid Outcome Prediction Model for Severe Traumatic Brain Injury. *J. Neurotrauma*, 24 (2007) 136-146
- [5] Rughani, A. I., Dumont, T. M., Lu, Z. et al.: Use of an Artificial Neural Network to Predict Head Injury Outcome. *J. Neurosurg.*, 113 (2010) 585-590
- [6] Brown, A. W., Malec, J. F., McClelland, R. L. et al.: Clinical Elements that Predict Outcome After Traumatic Brain Injury: A Prospective Multicenter Recursive Partitioning (Decision-Tree) Analysis. *J. Neurotrauma*, 22 (2005) 1040-1051
- [7] Rovlias, A., & Kotsou, S.: Classification and Regression Tree for Prediction of Outcome After Severe Head Injury using Simple Clinical and Laboratory Variables. *J. Neurotrauma*, 21 (2004) 886-893
- [8] Quinlan JR. Induction of Decision Trees. *Machine Learning*. Vol. 1, 1986, pp. 81-106.
- [9] Marcano-Cedeño A, Quintanilla-Domínguez J, Andina D. Breast cancer classification applying artificial metaplasticity algorithm. *Neurocomputing*. Vol. 38(8), 2011, pp. 9573-9579. doi:10.1016/j.neucom.2010.07.019
- [10] Specht DF. A general regression neural network. *IEEE Transactions on Neural Networks*. Vol. 2(6), 1991, pp. 568-576. doi:10.1109/72.97934
- [11] Denle D. Predicting breast cancer survivability: a comparison of three data mining methods. *Artificial Intelligence in Medicine*. Vol. 34(2), 2005, pp. 568-576. doi:10.1016/j.artmed.2004.07.002