
**EL PAPEL DEL SOFTWARE LIBRE EN EL PROCESO ENSEÑANZA-APRENDIZAJE
DEL RECONOCIMIENTO DE PATRONES**

**THE ROLE OF OPEN SOFTWARE IN THE TEACHING-LEARNING PROCESS ON
PATTERN RECOGNITION**

Cornelio Yáñez Márquez¹

cyanez@cic.ipn.mx

Yenny Villuendas Rey²

yenny@informatica.unica.cu

RESUMEN

El advenimiento de tecnologías novedosas ha estimulado una serie de cambios drásticos en la educación, que inducen un papel más activo e interactivo de los estudiantes. En este contexto, es posible mejorar de manera notable los procesos de enseñanza-aprendizaje en los cursos de posgraduados, mediante el uso de software libre y de repositorios de datos de acceso público. En este artículo se presentan algunas experiencias educativas que los autores han vivido en su grupo de investigación, cuando han usado herramientas de software libre y repositorios de datos de acceso público, en cursos de Reconocimiento de Patrones a nivel del posgrado.

ABSTRACT

The advent of novel technologies has encouraging education to undergo a series of drastic changes, inducing students to take a more active and interactive role. By coupling such facts with the use of open software and publicly accessible data repositories, the teaching-learning processes involved in postgraduate courses may be greatly enhanced. Some educative experiences in this regard, lived by the authors in their research group when using open software tools and public data sets repositories in the context of postgraduate level courses on Pattern Recognition, are presented in this paper.

¹ Licenciado en Física y Matemáticas, Máster en Ingeniería de Cómputo y Doctor en Ciencias de la Computación. Labora en el Centro de Investigación en Computación del Instituto Politécnico Nacional, México.

² Ingeniera en Informática y Máster en Informática Aplicada. Labora en el Departamento de Ciencias Informáticas de la Universidad de Ciego de Ávila, Cuba.

Palabras claves: Software libre, Proceso enseñanza-aprendizaje, Reconocimiento de Patrones.

Keywords: Open software, Teaching-learning process, Pattern recognition

INTRODUCCIÓN

En años recientes se ha observado una avalancha de avances tecnológicos que inciden en la actividad humana, influyen en cada acción de individuos y pueblos, y representan cambios drásticos en hábitos y actitudes. Las áreas de la computación y las comunicaciones han experimentado, especialmente, una sorprendente evolución, de modo que las computadoras digitales de la actualidad son dispositivos muy poderosos, capaces de realizar procesos impensables hasta hace pocos años, y de establecer enlaces entre los humanos sin importar las distancias (Bowdon & Carpenter, 2011).

Como una consecuencia positiva e inmediata del efecto que provocan estos avances en la vida cotidiana del ser humano, se observan cambios radicales en la educación del siglo XXI. Como atinadamente mencionan Webster y Murphy (2008), los ambientes educativos son testigos de la aparición constante de tecnologías novedosas, las cuales representan valiosas oportunidades para los procesos de enseñanza-aprendizaje. Según estos autores, entre el cúmulo de retos que surgen al aprender y enseñar con nuevas tecnologías se incluye el desarrollo de software abierto, herramientas para redes sociales, dispositivos móviles y sistemas de administración; es tarea de los educadores conseguir un balance entre estos factores que inciden de manera directa en los procesos educativos. Una conclusión natural a partir de estas consideraciones, es que las instituciones deben actuar estratégicamente a fin de fomentar la innovación y, al mismo tiempo, asegurar la calidad de la aplicación de los avances tecnológicos en los procesos de enseñanza y aprendizaje.

Por otro lado, como apunta Koranne (2011), las herramientas computacionales se han convertido en un elemento casi permanente que debe estar presente en los esquemas novedosos de aprendizaje interactivo, prácticamente para cada uno de los temas educativos en ciencias e ingeniería, incluyendo las ciencias de la computación y, específicamente, el reconocimiento de patrones. Entre todas las herramientas computacionales, el software libre juega un papel muy relevante, debido a las ventajas que su uso implica en los ambientes educativos. Además, y afortunadamente para los

investigadores, en la ciencia y la educación científica, una gran cantidad de datos producidos en los laboratorios de investigación científica a nivel mundial son almacenados y ofrecidos públicamente en diversos sitios web. Es un hecho indefectible que el uso de estos repositorios públicos de bancos de datos facilita la creación y diseño de modelos matemáticos para el análisis de datos y el reconocimiento de patrones (Robles et al., 2011).

En este artículo se presentan algunas experiencias educativas que los autores han vivido en su grupo de investigación, cuando han usado herramientas de software libre y repositorios de datos de acceso público, en cursos de Reconocimiento de Patrones a nivel de posgrado, tanto en México como en Cuba.

Se incluyen tres excelentes herramientas de software libre que facilitan significativamente los procedimientos inherentes al reconocimiento de patrones: Weka (Hall et al., 2009; Witten & Frank, 2005), MOA (Bifet, 2010) y Octave (Quarteroni & Saleri, 2006), las cuales se ilustran con casos de estudio; además, se presenta Moodle, una herramienta de desarrollo de contenido y manejo de plataformas (Antón-Rodríguez et al., 2012), se incluyen algunas reflexiones sobre las redes sociales y su impacto en los procesos educativos (Bicen & Uzunboylu, 2013), y se describen brevemente los repositorios de datos de acceso público, mencionando algunos ejemplos representativos.

Como colofón se muestran algunas experiencias educativas de los autores y sus grupos de investigación (Alfa-Beta y de Análisis de Datos), en relación con el uso de software libre y repositorios públicos de datos en los cursos de posgrado de Reconocimiento de Patrones.

DESARROLLO

Según afirman Brocco y Frapolli (2011), el propósito del uso de herramientas de software libre en los ámbitos educativos relacionados con las ciencias de la computación, es doble: por un lado, al estudiante se le proporcionan herramientas libres para crear, modificar e intercambiar información; y por otro, se permite al profesor explicar los fundamentos de las ciencias de la computación en escenarios reales.

Dado que el reconocimiento de patrones es una importante rama de las ciencias de la computación, las opiniones previas se aplican, sin cambios, al contenido del presente

artículo; aún más, como extensamente explica Janert (2011), es posible, muy práctico y poco oneroso realizar tareas de reconocimiento de patrones con herramientas de software libre.

La plataforma Weka

Weka es una plataforma de software libre que contiene una colección de algoritmos (de machine learning) que realizan tareas de reconocimiento de patrones y minería de datos, tales como pre-procesamiento, clasificación, regresión, agrupamiento, reglas de asociación y visualización de datos. Weka contiene facilidades para desarrollar esquemas nuevos de reconocimiento de patrones (Hall et al., 2009).

Weka es el acrónimo de “*Waikato Environment for Knowledge Analysis*”. Esta joya de software libre fue desarrollada en la Universidad de Waikato, Nueva Zelanda, por un equipo de investigadores dirigidos por Ian H. Witten y Eibe Frank (Witten & Frank, 2005). El software está escrito en Java, lo que garantiza su portabilidad, de modo que se ejecuta en Windows, Linux, and Mac. Weka se distribuye bajo la licencia GNU y puede obtenerse libremente del sitio <http://www.cs.waikato.ac.nz/ml/weka/>

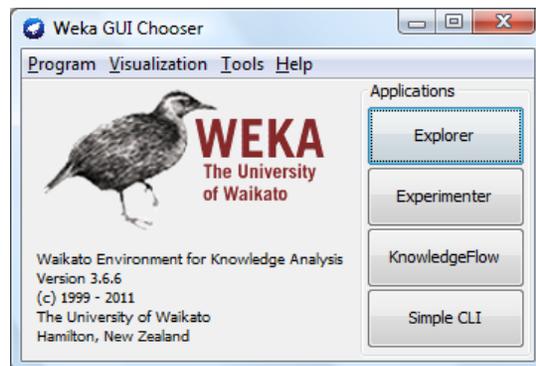


Fig. 1. Pantalla principal de Weka.

Como se ilustra en la figura 1, Weka se compone de cuatro aplicaciones principales: *Explorer*, *Experimenter*, *Knowledge Flow*, y *Simple CLI*, las cuales apoyan a los investigadores en la realización de todas las tareas relevantes del reconocimiento de patrones y la minería de datos. En nuestro grupo de investigación se usa Weka de manera cotidiana en artículos y tesis, tanto por investigadores como por estudiantes.

La plataforma MOA

La llamada “computación verde” es el estudio y práctica de usar eficientemente los recursos computacionales; los ahorros de estos recursos se manifiestan, normalmente,

como eficiencia algorítmica. Un ámbito que demanda este tipo de computación, es el relacionado con los procesos de *knowledge discovery from data streams*, donde ya no es posible aplicar algoritmos tradicionales que analizan conjuntos de datos estáticos, para los que se asumen distribuciones estacionarias y memoria ilimitada; en los *data streams*, por el contrario, los datos fluyen a altas velocidades, y los algoritmos que se requieren para procesar estos flujos de datos exhiben restricciones severas de espacio y tiempo (Bifet, 2010).

MOA (*Massive Data Analysis*) es la respuesta. MOA es una plataforma de software libre para aprendizaje en línea a partir de *data streams*, la cual contiene una colección de algoritmos (de *machine learning*) que realizan tareas de reconocimiento de patrones y minería de datos en flujos de datos.

Al igual que Weka, el desarrollo de MOA se realizó en la Universidad de Waikato, Nueva Zelanda: en esta ocasión, el equipo de investigadores fue dirigido por Geoff Holmes y Bernhard Pfahringer. El software también está escrito en Java, lo que garantiza su portabilidad. MOA se distribuye libremente y puede obtenerse del sitio <http://sourceforge.net/projects/moa-datastream/files/MOA/2013%20August/moa-release-2013.08.zip/download>.

La plataforma Octave

Octave es un lenguaje de programación interpretado de alto nivel, diseñado para cómputo numérico; proporciona facilidades para la solución numérica de problemas lineales y no lineales, y para ejecutar una gran variedad de experimentos numéricos. También incluye amplias capacidades gráficas para visualización y manipulación, por lo que se convierte en la alternativa de software libre de Matlab; así, al igual que este software comercial, Octave es una plataforma de programación que permite el manejo fácil de vectores y matrices, además de brindar ambientes de visualización en dos y tres dimensiones (Quarteroni & Saleri, 2006).

La plataforma Octave fue creada y desarrollada por James B. Rawlings de la Universidad de Wisconsin-Madison y John G. Ekerdt de la Universidad de Texas; este software libre está escrito en C y C++, se distribuye libremente bajo la licencia GNU, y está disponible en este sitio: <http://www.octave.org/download.html>.

La plataforma Moodle

La página principal del sitio web moodle.org contiene una explicación muy clara y concisa de los motivos que llevaron a la creación de la plataforma de software libre Moodle. Ahí se puede leer textualmente: “*Welcome to the Moodle community! Moodle is a Course Management System (CMS), also known as a Learning Management System (LMS) or a Virtual Learning Environment (VLE). It is a free web application that educators can use to create effective online learning sites. Moodle.org is our community site where Moodle is made and discussed.*”

La piedra angular pedagógica de Moodle es la filosofía del aprendizaje constructivista, el cual está basado en la forma en que los seres humanos construyen su conocimiento, mediante la interacción con los demás seres humanos y con los materiales de aprendizaje (Antón-Rodríguez et al., 2012). En este sentido, Moodle permite que el instructor integre en un curso diferentes recursos, tales como materiales estáticos de aprendizaje, actividades interactivas o incluso situaciones de tipo social, ayudando así a los estudiantes a fomentar sus potencialidades de aprendizaje.

Redes sociales

La brutal e intempestiva incursión de las redes sociales en la vida cotidiana de los seres humanos ha venido a propiciar profundos cambios en el mundo contemporáneo. En lo político, baste recordar el papel fundamental que jugaron estas tecnologías de vanguardia en la primavera árabe, nombre que se ha acuñado para designar un conjunto de movimientos sociales que se suscitaron en varios países árabes a partir de 2010. A su vez, la primavera árabe sirvió de inspiración a otras manifestaciones político-sociales no menos importantes, como el movimiento 15M (movimiento de los indignados) que se inició en España en 2011 y permeó a otros países de varios continentes; del *Occupy Wall Street*, que se extendió de manera masiva a países de América Latina y Europa; y de los recientes y sorprendentes movimientos masivos en Brasil, en el marco del fútbol. En todos estos relevantes acontecimientos, han jugado un papel preponderante las redes sociales, particularmente el Twitter y el Facebook, los cuales prácticamente han invalidado la censura de los medios oficiales.

Pero esta activa preponderancia de las redes sociales no se da solo en la manifestación de las ideas políticas que luchan por justicia y democracia; también se manifiesta de manera ostensible en los procesos educativos, en todas las modalidades y niveles. Por

ejemplo, Bicen y Uzunboylu (2013) utilizaron Facebook y Web 2.0 en la investigación educativa, creando un efecto positivo cuando la usaron en ambientes de aprendizaje en línea; su opinión es que la educación en línea y las tecnologías integradas a la educación son ahora indispensables, dado que la comunicación es más fácil en nuestros días gracias a la tecnología. Con el uso de internet, algunas actividades casi imposibles de realizar en el salón de clases, pueden ahora realizarse fácilmente con el uso de teléfonos inteligentes. Con el creciente uso de las redes sociales, se incrementan también las necesidades de comunicación y de intercambio de información entre la gente; los autores concluyen que gracias a las redes sociales, los estudiantes pueden tener una educación de más alto nivel.

Otro importante ejemplo del uso de las redes sociales en la educación, es un extenso estudio que proporciona una revisión crítica del uso de Twitter en los ámbitos educativos (Dhir et al., 2013). Después de una profunda revisión bibliográfica, los autores de este estudio discutieron los posibles beneficios y problemas que produce el Twitter en la educación, y concluyeron que esta herramienta de redes sociales tiene un impacto positivo en la educación informal, la dinámica de las clases y la motivación.

Repositorios de acceso público

El UCI *Machine Learning Repository* es el repositorio de datos de acceso público por antonomasia; es el más famoso y uno de los más utilizados por estudiantes, educadores e investigadores de todo el mundo. Según se informa en el sitio web donde está alojado, <http://archive.ics.uci.edu/ml/about.html>, este repositorio es una colección de bases de datos y de generadores de datos que son usados por la comunidad de *machine learning* y disciplinas afines, como lo es el reconocimiento de patrones, para el análisis empírico de algoritmos. En el sitio mismo se da a conocer un dato que exhibe el impacto del repositorio: ha sido citado más de mil veces, convirtiéndolo en uno de los 100 “artículos” más citados en ciencias de la computación.

Es tal la penetración que ha tenido este estilo de facilitar datos a los investigadores, que existe un repositorio de datos de acceso público casi para cualquier disciplina científica; por ejemplo el *Protein Data Bank* (PDB) es un repositorio de acceso público de datos estructurales de moléculas biológicas grandes, tales como proteínas y ácidos nucleicos, en 3D (www.rcsb.org/pdb/).

Algunas asociaciones internacionales cuya actividad está relacionada con bancos de datos también hacen públicos sus repositorios de bancos de datos, a fin de que científicos y estudiantes de todo el mundo los usen en sus trabajos de investigación. En este sentido, la *International Association for Pattern Recognition* (IAPR) es una asociación internacional sin fines de lucro, que aglutina a organizaciones científicas o profesionales de todo el mundo, las cuales estén relacionadas con reconocimiento de patrones, visión por computadora y procesamiento de imágenes en el más amplio sentido.

Entre los múltiples comités técnicos de la IAPR, hay uno en particular, el IAPR TC 5 (*Benchmarking & Software*), cuyo propósito es proporcionar recursos para asistir a los investigadores en la implementación de *benchmarks* para sistemas de reconocimiento de patrones. El sitio web del TC 5 contiene bancos públicos de datos relacionados con una gran variedad de tópicos: hay bancos de datos para reconocimiento de símbolos escritos a mano, secuencias de video del proyecto CAVIAR (*Context Aware Vision using Image-based Active Recognition*), datos del instituto Max Planck, y hasta un generador de bancos de datos.

El sitio web <http://www.chemweb.com/content/databases> contiene una gran cantidad de bases de datos para química, mientras que hay sitios web que más que repositorios son meta-repositorios, como el sitio <http://www.kdnuggets.com/datasets/index.html>, el cual contiene bases de datos de repositorios de datos de acceso público.

La importancia de los repositorios de datos de acceso público es tal, que autores como Marco (2000) han publicado explicaciones muy detalladas sobre el uso eficiente de estos repositorios.

La cantidad y calidad de los repositorios de datos de acceso público crece continuamente, provocando que existan repositorios públicos para cada necesidad, en áreas como ciencias de la computación, reconocimiento de patrones, redes neuronales, aprendizaje supervisado, memorias asociativas, inteligencia artificial, inteligencia computacional y cómputo no convencional, por citar solo algunas.

Experiencias educativas en nuestro grupo de investigación

En todos los casos, la estrategia general de trabajo consta de pasos similares, a saber: primeramente, cada equipo de investigación formado por estudiantes y profesores de

los grupos de investigación Alfa-Beta y de Análisis de Datos, escoge el problema a resolver; acto seguido, se adquieren los bancos de datos del repositorio de datos de acceso público más apropiado; después, se desarrolla e implementa, en alguna herramienta de software libre, un modelo original de reconocimiento de patrones, el cual estará basado en redes neuronales, memorias asociativas, algoritmos meta-heurísticos, teoría de testores, conjuntos compactos, o una combinación de estas; y finalmente, se usa Weka para probar el modelo y realizar un estudio comparativo de rendimiento comparando, para ello, los resultados arrojados por el modelo con los obtenidos al utilizar los modelos clásicos del estado del arte sobre el mismo banco de datos.

- **Predicción inteligente de contaminantes atmosféricos**

Aquí presentamos el primer caso de estudio relacionado con las experiencias educativas en nuestro grupo de investigación; se trata de seguir la estrategia descrita, a fin de realizar predicciones inteligentes de la concentración de contaminantes ambientales en la Ciudad de México (López-Yáñez et al., 2011). Los datos provienen de la Red Automática de Monitoreo Atmosférico (RAMA), que es un banco de datos del repositorio de acceso público del Sistema de Monitoreo Atmosférico de la Ciudad de México (SIMAT), y el modelo de predicción usado es una adaptación del Clasificador Gamma para la predicción de series de tiempo.

Para construir el conjunto de aprendizaje, se tomaron muestras de la concentración de CO (en ppm), en una estación específica (Iztacalco, para este caso), durante todo el año 2010; luego, se concatenaron las muestras de 10 en 10 (avanzando una muestra en cada paso) para formar un conjunto de 8749 de patrones de aprendizaje, con la muestra siguiente a cada patrón de entrada como el correspondiente patrón de salida. Estas 8749 asociaciones de patrones se usan para la fase de aprendizaje del Clasificador Gamma, el cual se implementó en Octave. El conjunto de prueba, formado por 661 asociaciones, se construyó de manera similar a como se hizo con el conjunto de aprendizaje, tomando datos en la misma estación, pero de febrero de 2011.

Las predicciones de CO son de gran importancia al calcular las predicciones del IMECA (Índice Metropolitano de la Calidad del Aire). Por ejemplo, a las 19:00 hrs. del día 2 de febrero, el sistema propuesto predijo una concentración de CO de 0.8 ppm, siendo precisamente ese el valor real: 0.8ppm; como era de esperarse, los correspondientes

valores del IMECA son iguales (con valor 7), el cual equivale a condición Buena. Por otro lado, a las 13:00 hrs. del día 16 de febrero, los valores predicho y real son 1.7 ppm y 1.4 ppm, respectivamente, mientras los correspondientes valores del IMECA son 15 y 13, los cuales equivalen, ambos, a una condición Buena; así, aunque la predicción del CO no fue exacta, el valor del IMECA se predijo correctamente. Finalmente, a las 6:00 hrs. del día 23 de febrero, el valor predicho para la concentración de CO en esa estación fue de 0.6 ppm, mientras que el valor real fue de 1.8 ppm, arrojando así los valores 5 y 16 para el IMECA, siendo Buena la condición en ambos casos.

Usando los mismos conjuntos de datos, se usó Weka para realizar un estudio comparativo de rendimiento del Clasificador Gamma, contra los resultados arrojados por los más de 60 modelos incluidos en Weka. El resultado es notable: el Clasificador Gamma exhibe el segundo mejor rendimiento para la predicción de concentraciones de CO, siendo solo superado por SMOreg (SVM *based regression*), el cual es un modelo de predicción altamente sofisticado y computacionalmente caro.

- **Elaboración inteligente de diagnósticos médicos**

En este segundo caso de estudio se sigue la estrategia descrita, con el propósito de elaborar diagnósticos médicos de manera inteligente, aprovechando para ello, de manera intensiva, los bancos de datos del UCI *Machine Learning Repository* y las bondades de la plataforma Weka. En este caso se creó un nuevo modelo de clasificación inteligente de patrones, AMBC, que es el acrónimo de *Associative Memory based Classifier* (Aldape-Pérez et al., 2012).

Durante la fase experimental, se diagnosticaron siete enfermedades con el AMBC tomados del UCI *Machine Learning Repository*, y se comparó su rendimiento con el arrojado por Weka para los más de 60 algoritmos de clasificación que se incluyen en esa plataforma, de los cuales se tomaron los mejores veinte para fines del estudio comparativo. Resultado: el modelo AMBC arrojó el mejor rendimiento, en promedio, respecto a los siete bancos de datos.

Los siete bancos de datos que se usaron son: *Haberman survival dataset* (sobrevivencia de pacientes que han experimentado cirugía de cáncer de seno), *Liver disorders dataset* (contiene desórdenes ocasionados por el consumo excesivo de alcohol), *Acute inflammations dataset* (pacientes con enfermedades del sistema

urinario), *Pima Indians diabetes dataset* (pacientes aquejados de diabetes), *Breast cancer dataset* (pacientes con cáncer de mamas), *Heart disease dataset* (pacientes con trastornos cardíacos) y *Hepatitis disease dataset* (pacientes aquejados de hepatitis).

Los resultados experimentales indican que ninguno de los algoritmos supera a los demás en todas las enfermedades; sin embargo, cabe hacer notar que el AMBC consiguió los mejores resultados en tres de los siete bancos de datos al aplicar el método de validación hold-out con las particiones 50-50 y 70-30. Además, el AMBC logró los mejores resultados en cuatro de los siete bancos de datos al aplicar el método de validación *ten-fold cross validation*.

- **Predicción inteligente del esfuerzo en el desarrollo de proyectos de software**

El tercer caso de estudio trata sobre el uso del Clasificador Gamma en la predicción inteligente del esfuerzo desplegado en el desarrollo de proyectos de software (López-Martín et al., 2012). Al respecto es preciso hacer notar que, típicamente, los gerentes de software basan sus predicciones en juicios personales, a fin de estimar el esfuerzo, el cual se traduce en el número de horas que se invertirán en el desarrollo de un proyecto. El Clasificador Gamma se entrenó con un conjunto de datos de 163 proyectos de software, mientras que en la fase experimental se usaron otros 68 proyectos.

Los resultados experimentales mostraron que el Clasificador Gamma superó, al menos, a dos de las técnicas más usadas en este tipo de problemas: la lógica difusa y la regresión estadística.

- **Predicción inteligente del grupo de orientación pedagógica de familias de menores con trastornos de la conducta**

El cuarto caso de estudio trata sobre el uso de herramientas de clasificación supervisada para la determinación del grupo de orientación pedagógica, al que deben ser asignadas las familias de los menores con trastornos de la conducta en Cuba. La atención a los menores con este tipo de trastornos es una prioridad del sistema de educación; dentro de esta, la orientación adecuada a las familias de los menores es crucial. Al respecto, es preciso hacer notar que cada tipo de familia necesita de una orientación personalizada, de acuerdo con sus necesidades. Para dar solución a esta problemática, se utilizó un clasificador del vecino más cercano, con una función de similitud especialmente diseñada para la comparación de descripciones de familias.

También se aplicaron técnicas de pre-procesamiento de datos, con el fin de seleccionar las familias más relevantes, y utilizarlas como prototipos en el proceso de clasificación. El clasificador se entrenó con un conjunto de datos de las familias de menores con trastornos de la conducta de Ciego de Ávila, hasta el año 2010, usando en la fase experimental el procedimiento de validación cruzada (k-fold cross validation).

Los resultados experimentales mostraron que el pre-procesamiento de los datos fue adecuado, ya que se logró un 100% de clasificaciones correctas en la fase de prueba.

- **Detección de niños de edad preescolar con altas potencialidades para el desarrollo**

El quinto caso de estudio trata sobre la detección temprana de altas potencialidades para el desarrollo en niños cubanos del grado preescolar. Para ello, se utilizaron diferentes indicadores que apuntan hacia la presencia de altas potencialidades en los niños de esta edad. Se utilizaron elementos de la teoría de testores, y conjuntos compactos, para lograr un adecuado pre-procesamiento de los datos referentes a los niños. Se entrenó el clasificador k-NN y se utilizó en la detección inteligente de la presencia de altas potencialidades en los niños, utilizando una función de similitud diseñada especialmente para este problema. Al respecto es preciso hacer notar que, típicamente, las educadoras del grado preescolar se basan en su experiencia para determinar si un niño tiene o no altas potencialidades, y no siempre estas son detectadas de forma temprana.

El clasificador NN se entrenó con datos pre-procesados, correspondientes a 4 aulas del grado preescolar de Ciego de Ávila. Los resultados experimentales mostraron un 94% de detecciones correctas en la fase de prueba, superando a otras técnicas, como árboles de decisión y redes neuronales artificiales.

CONCLUSIONES

En este artículo se han descrito experiencias educativas y de investigación que se han experimentado en el seno de los grupos de investigación Alfa-Beta y de Análisis de Datos, en relación con el uso de herramientas de software libre como Weka, Octave y MOA, y repositorios de datos de acceso público (tales como el UCI *Machine Learning Repository*). Todo lo anterior, en el contexto de cursos de posgrado en Reconocimiento

de Patrones.

Específicamente, profesores y estudiantes, tanto de México como de Cuba, han empleado estas herramientas y repositorios en el afán de mejorar los procesos de enseñanza y aprendizaje en los cursos de posgrado y en su formación científica en general, con énfasis en los cursos relacionados con el Reconocimiento de Patrones.

REFERENCIAS BIBLIOGRÁFICAS

- Aldape-Pérez, M., Yáñez-Márquez, C., Camacho-Nieto, O., & Argüelles-Cruz, A.J. (2012). An associative memory approach to medical decision support systems. *Computer Methods and Programs in Biomedicine* 103, 3, 287-307.
- Antón-Rodríguez, M., Pérez-Juárez, M.A., Díaz-Pernas, F.J., Perozo-Rondón, F.J., Martínez-Zarzuela, M., & González-Ortega, D. (2012). Moodle-Based Software to Support the Learning of Web Programming. *International Journal of Knowledge Society Research* 3, 3, 16-28.
- Bicen, H., & Uzunboylu, H. (2013). The Use of Social Networking Sites in Education: A Case Study of Facebook. *Journal of Universal Computer Science* 19, 5, 658-671.
- Bifet, A. (2010). *Adaptive Stream Mining: Pattern Learning and Mining from Evolving Data Streams*. Amsterdam, Netherlands: IOS Press BV.
- Bowdon, M.A. & Carpenter, R.G. (2011). *Higher Education, Emerging Technologies, & Community Partnerships*. Hershey, PA, USA: Information Science Reference.
- Brocco, A., & Frapolli, F. (2011). *Open Source in Higher Education: Case Study Computer Science at the University of Fribourg*. Diploma of Advanced Studies, Centre de Didactique Universitaire, University of Fribourg, Switzerland.
- Dhir, A., Buragga, K., & Boreqqah, A.A. (2013). Tweepers on Campus: Twitter a Learning Tool in Classroom?. *Journal of Universal Computer Science* 19, 5, 672-691
- Hall, M., Frank, E., Holmes, G., Pfahringer, B., Reutemann, P., & Witten, I.H. (2009).

- The Weka Data Mining Software: An Update. *SIGKDD Explorations*, 11, 1, 10-18.
- Janert, P.K. (2011). *Data Analysis with Open Source Tools*. Sebastopol, CA. O'Reilly Media, Inc.
- Koranne, S. (2011). *Handbook of Open Source Tools*. New York, NY: Springer+BusinessMedia.
- López-Martín, C., López-Yáñez, I., & Yáñez-Márquez, C. (2012). Application of Gamma Classifier to Development Effort Prediction of Software Projects. *Applied Mathematics & Information Sciences* 6, 3, 411-418.
- López-Yáñez, I., Argüelles-Cruz, A.J., Camacho-Nieto, O., & Yáñez-Márquez, C. (2011). Pollutants time series prediction using the Gamma classifier. *International Journal of Computational Intelligence Systems* 4, 4, 680-711.
- Marco, D. (2000). *Building & Managing the Metadata Repository: A Full Lifecycle Guide*. New York, NY: John Wiley & Sons, Inc.
- Quarteroni, A., & Saleri, F. (2006). *Scientific Computing with MATLAB & Octave*. Berlin/Heidelberg: Springer-Verlag.
- Robles, G., Gonzalez-Barahona, J.M., Izquierdo-Cortazar, D., & Herraiz, I. (2011). Tools & datasets for mining libre software repositories. In Stefan, K. (Ed.), *Multi-Disciplinary Advancement in Open Source Software & Processes* (pp. 24-42). Hershey, PA, USA: IGI Global.
- Webster, L. & Murphy, D. (2008). Enhancing Learning Through Technology: Challenges & Responses. In Kwan, R., Fox, R., Chan, F.T., & Tsang, P. (Eds.), *Enhancing Learning Through Technology: Research on Emerging Technologies & Pedagogies* (pp. 1-16). Singapore: World Scientific.
- Witten, I.H., & Frank, E. (2005). *Data mining: practical machine learning tools & techniques*. Amsterdam, Netherlands: Elsevier.

Recibido: diciembre de 2013
Aceptado para su publicación: marzo de 2014