

## Algoritmo genético para la generación de reglas de predicción en los cursos de moodle.



*Genetic algorithm for the generation of prediction rules in moodle courses.*

Dr. Yasser Vázquez Alfonso PhD.<sup>1</sup>, MSc. Amaury Machín Armentero<sup>2</sup>, MSc. Velastegui López Luis Efraín<sup>3</sup>, Ing. Víctor Peñafiel Gaibor PhD.<sup>4</sup>

Recibido: 13-09-2017 / Revisado: 07-11-2017 Aceptado: 05-12-2017/ Publicado: 01-01-2018

### Abstract.

DOI: <https://doi.org/10.33262/cienciadigital.v2i1>.

The development of new technologies of education and communication has made possible the use of the Internet and more specifically the World Wide Web in distance education, giving rise to the denonimada Web-based Education or elearning. This implies a new conception of education by incorporating new technologies into the teaching and learning process. The objective of the research was to apply a Genetic Algorithm for the analysis of how students learn on the web, starting from their navigation patterns and making use of virtual platforms in the university context. The proposed algorithm allowed optimizing and customizing the structure of Moodle courses. administrative models that manage the informational asset, and allow proposing actions, programs and projects to implement business intelligence systems.

**Keywords:** Genetic Algorithms, Moodle, Navigation Patterns.

### Resumen.

El desarrollo de las nuevas tecnologías de la educación y la comunicación han hecho posible la utilización de Internet y más concretamente la WWW (World Wide Web) en la educación a distancia, dando lugar a la denominada Educación basada en Web o e-learning. Esto supone una nueva concepción de la educación mediante la incorporación de nuevas tecnologías al proceso de enseñanza y aprendizaje. En la

<sup>1</sup> Universidad de la Habana, Cuba, yasser@fur.uh.cu

<sup>2</sup> Universidad Agraria de la Habana, Cuba, amaurym@unah.edu.cu

<sup>3</sup> Editorial Ciencia Digital, Ambato, Ecuador, luisefrainvelastegui@cienciadigital.org

<sup>4</sup> Universidad Técnica de Ambato, Facultad de Ciencias de la Salud, Ambato, Ecuador, vi.penafiel@uta.edu.ec

investigación realizada se propuso como objetivo aplicar un Algoritmo Genético para el análisis de como los estudiantes aprenden en el web, partiendo de sus patrones de navegación y haciendo uso de las plataformas virtuales en el contexto universitario. El algoritmo propuesto permitió optimizar y personalizar la estructura de los cursos del Moodle.

**Palabras Claves:** Algoritmos Genéticos, Moodle, patrones de navegación.

## **Introducción.**

El desarrollo de las nuevas tecnologías de la educación y la comunicación han hecho posible la utilización de Internet y más concretamente la WWW (World Wide Web) en la educación a distancia, dando lugar a la denominada Educación basada en Web o e-learning (Horton, 2000). En los últimos años, e-learning supone una nueva concepción de la educación mediante la incorporación de nuevas tecnologías al proceso de enseñanza y aprendizaje. Esta plataforma educativa proporciona tanto a docentes como alumnos una importante variedad de recursos, lo que en muchas ocasiones esto suele ser una desventaja.

Esto hace que el alumno se encuentre desorientando frente a la elección de recursos diseñados. Este problema requiere una reestructuración de cómo el docente ha diseñado la asignatura, con el objetivo que ese diseño se adapte mejor manera a las características e intereses de los alumnos (Corso and Alfaro, 2007).

El uso de ambientes virtuales basados en la web, como herramienta de apoyo al proceso educativo, se ha expandido en los últimos años, gracias al desarrollo de nuevos enfoques y modalidades de educación, como el e-learning y el b-learning. Estas nuevas formas de utilización de la tecnología como apoyo al proceso de enseñanza aprendizaje, han incentivado la aparición de software especializado en proveer los mecanismos que facilitan el desarrollo del trabajo autónomo del estudiante, su interacción con el docente, y el uso de herramientas colaborativas y de comunicación para el trabajo grupal (Aponte et al., 2012).

Estas técnicas de minería de datos o data mining (DM) ya se han aplicado con éxito en sistemas de comercio electrónico o e-commerce, para comprender el comportamiento de clientes en línea de sistemas de comercio electrónico y poder incrementar las ventas (Srivastava et al., 2000). Para conseguir su objetivo, las herramientas de DM utilizan técnicas de extracción de conocimiento para descubrir información útil en la mejora del sistema. Aunque los métodos de descubrimiento de información utilizados en ambas áreas (e-commerce y elearning) son similares, los objetivos finales tienen matices totalmente diferentes debido a que en e-commerce el objetivo es guiar a los clientes durante la compra para maximizarla, mientras que en e-learning el objetivo es guiar a los estudiantes durante

su aprendizaje para maximizarlo. Por lo tanto, cada uno tiene unas características específicas que requieren de un tratamiento diferente dentro del problema de minería de Web.

En esta investigación se presenta un algoritmo genético cuyo objetivo es obtener un conjunto de reglas de asociación que permitan la realización de predicciones en cuanto al comportamiento de los usuarios y aporten conocimiento interesante sobre un problema con relaciones desconocidas entre las variables. El algoritmo tiene capacidad para extraer reglas difusas o nítidas en función de que las variables del problema sean continuas o nominales. Estas reglas pueden permitir extraer conclusiones y planificar de mejor manera la organización de los cursos en Moodle.

## **2-Desarrollo.**

En los últimos años, el aprendizaje electrónico o conocido también por el anglicismo, e-learning, ha generado una nueva concepción de la educación mediante la incorporación de nuevas tecnologías al proceso de enseñanza y aprendizaje. El e-learning ha posibilitado que los estudiantes puedan comunicarse y colaborar con otros estudiantes y docentes (profesores, tutores, mentores, entre otros), de forma síncrona o asíncrona, sin limitaciones espacio-temporales. En otras palabras, el e-learning se comporta como una modalidad de aprendizaje dentro de la educación a distancia en la que se utilizan las redes de datos como medios (internet, intranets, entre otros), las herramientas o aplicaciones hipertextuales como soporte (correo electrónico, web, chat, entre otros) y los contenidos y/o unidades de aprendizaje en línea como materiales formativos (desde simples imágenes, audio, video, documentos, entre otros, hasta complejas producciones multimedia, así como contenidos contruidos de forma colaborativa), entre otros (García et al., 2008).

El aumento de la utilización de las plataformas e-learning ha generado en los últimos años una fuerte tendencia por parte de las universidades e institutos de enseñanza en el uso de plataformas educativas basadas en la web, centrando su uso para cursos a distancia, semipresenciales y presenciales. Según (Palma, 2013) el uso de estas plataformas virtuales han posibilitado además, la incorporación de herramientas tecnológicas en las aulas de clases, lo cual constituye un enriquecimiento del currículo educativo en el cual, el docente pasa de cumplir sólo un rol de transmisor de conocimientos a ser un facilitador del proceso enseñanza – aprendizaje, dejando de constituir la fuente principal de información.

Al realizar una navegación por la web, los usuarios dejan huellas digitales, tales como direcciones de Protocolo de Internet o dirección IP (Internet Protocol por sus siglas en inglés), navegador empleado, cookies, url de los sitios visitados, entre otros datos que los servidores almacenan automáticamente en una bitácora de accesos (logs). El análisis de ficheros logs de los servidores web puede proporcionar variedad de información valiosa

que permitiría mejorar la estructura de un sitio web, así como permite estudiar el comportamiento de los usuarios al visitar estos sitios con el objetivo principal de crear una navegación más efectiva y un acceso más eficiente.

Para el análisis de estos ficheros log en el siguiente artículo se propone la aplicación de un algoritmo genético (Agrawal and Srikant) multiobjetivo para la extracción de reglas de asociación.

## **2-1. Reglas de Predicción.**

El modelado de dependencias (Klösgen and Zytkow, 2002), también denominado por algunos autores como inducción de reglas de predicción o reglas generalizadas tiene como cuyo objetivo el descubrimiento de reglas interesantes para mostrárselas al usuario. Estas reglas, que representan relaciones de dependencia importantes entre los datos y que se pueden utilizar para la posterior toma de decisiones, presentan el siguiente formato:

### **SI Cond1, Y...Condi... Y Condm ENTONCES Pred.**

Donde cada condición Condi y la predicción de la regla Pred están formados por una tripleta: (Atributo, Operador, Valor). Este tipo de reglas muestra la relación existente entre el antecedente, que contiene las condiciones sobre los valores de los atributos predictores, y el consecuente, que contiene la predicción sobre el valor del atributo objetivo.

Llegado este punto, habría que puntualizar la diferencia que existe entre el descubrimiento de reglas de predicción y el descubrimiento de reglas de asociación (Freitas, 2000), una tarea similar pero algo más general, en la que el objetivo es la búsqueda de todas las posibles relaciones entre atributos y donde puede haber incluso varios atributos en el consecuente de la regla. La tarea de descubrimiento de reglas de asociación, introducida por Agrawal (Agrawal and Srikant, 1994) que la define como el problema de encontrar todos los elementos que son frecuentes con respecto a un umbral mínimo de soporte y confianza.

El soporte indica el porcentaje de instancias que contienen tanto consecuente como el antecedente y la confianza indica el porcentaje de instancias que contienen el consecuente también contienen al antecedente. De forma, que el usuario debe especificar un valor mínimo de la medida de soporte y confianza, siendo el objetivo de la tarea encontrar todas las reglas que superen esos valores. Una regla de predicción puede verse como una regla de asociación con un solo elemento en su consecuente, por lo que cualquier algoritmo de descubrimiento de reglas de asociación se puede modificar fácilmente para añadir esta restricción y descubrir sólo reglas de predicción. Hemos utilizado reglas de predicción en lugar de reglas de asociación debido a que se pueden utilizar más fácilmente para la toma

de decisiones, son intuitivamente más comprensibles y muestran relaciones directas entre los elementos en lugar de todas las posibles relaciones.

## **2-2. Descubrimiento de Reglas de Predicción utilizando Algoritmos Evolutivos.**

La tarea del descubrimiento de reglas ha sido abordada desde multitud de paradigmas: construcción de árboles de decisión, aprendizaje inductivo, aprendizaje basado en instancias y, más recientemente redes neuronales y algoritmos evolutivos (Freitas, 2000).

El tipo de búsqueda que realizan cada uno de estos algoritmos va a determinar dónde se encuentran localizados dentro del panorama de la minería de reglas y desde el punto de vista de la minuciosidad de la búsqueda.

Los Algoritmos Evolutivos son algoritmos estocásticos de búsqueda basados en las ideas de la evolución darwiniana. Los paradigmas de Computación Evolutiva que se han aplicado para resolver el problema del descubrimiento de reglas (Freitas, 2013) son los Algoritmos Genéticos y la Programación Genética.

La Programación Genética (PG) se puede considerar como un paradigma de búsqueda más abierta que el de Algoritmos Genéticos.

La búsqueda realizada por la PG puede ser muy útil para clasificación y otras tareas, ya que el sistema puede producir diferentes combinaciones de atributos, utilizando las funciones disponibles en un conjunto preestablecido por la codificación, que no se considerarían utilizando un algoritmo genético convencional.

La Programación Genética basada en gramáticas (Romero et al., 2005b) es un paradigma de programación genética en el que los individuos vienen representados como árboles de derivación de una gramática definida por el usuario para especificar el espacio de soluciones al problema.

Se ha elegido este paradigma por la expresividad que presenta, que va a facilitar enormemente la interacción con el usuario. A continuación vamos a describir el algoritmo evolutivo utilizado y la codificación y la función de evaluación concreta de los individuos.

### **Algoritmo genético.**

El algoritmo evolutivo que hemos empleado para realizar la tarea específica de descubrimiento de reglas de predicción (Romero et al., 2005b), se muestra en la Figura 1.1.

**Figura 1.1:** Algoritmo evolutivo utilizado para la búsqueda de reglas de predicción.



**Fuente:** (Romero et al., 2005a).

El algoritmo comienza con la inicialización de la población, consiste en la generación del conjunto o población inicial de individuos o reglas. A continuación se realiza la evaluación, que consiste en calcular el ajuste de cada individuo y el almacenamiento de los mejores (en nuestro caso los no dominados) en una población final. Después se determina si el algoritmo debe de terminar, es decir, si ha alcanzado un número determinado de evoluciones o ha encontrado un número determinado de reglas. Sino, se pasa a la etapa de selección donde se elige de entre la población actual y la final a los individuos que van a ser padres de la siguiente etapa de reproducción. La reproducción consiste en la creación a partir de los padres seleccionados de nuevos individuos mediante los operadores de cruce y mutación (Michalewicz, 1996) .Finalmente la población actual es sustituida por la nueva población de padres y el proceso se vuelve a repetir.

### **Función de Evaluación.**

La función de evaluación mide la calidad de los individuos o reglas en nuestro caso. En la bibliografía hay descritas una gran cantidad de métricas para evaluar reglas (Lavrač et al., 1999) (Tan and Kumar, 2000) (soporte, confianza, interés, precisión, informatividad, fiabilidad negativa, sensibilidad, especificidad, cobertura, innovación, satisfacción, precisión relativa, etc.). Pero cada una mide un aspecto de la regla. Este problema sugiere el uso de una aproximación multiobjetivo (Fonseca and Fleming, 1993) para el descubrimiento de reglas, donde el valor de la función ajuste a optimizar no es un valor escalar único, sino un vector de valores, donde cada valor mide un aspecto diferente de la calidad de la regla. En nuestro caso la función de evaluación o función de ajuste utilizada está formada por un vector de tres componentes donde cada uno mida uno de los siguientes criterios de los individuos:

**Exactitud de la regla.** Mide la exactitud o precisión de las reglas. Nosotros hemos utilizado la medida denominada factor de certeza (Shortliffe and Buchanan, 1975).

**Comprensibilidad de la regla.** Mide la comprensibilidad de la regla por parte del usuario. Nosotros hemos utilizado la medida de simplicidad (Liu and Kwok, 2000) que depende de la longitud de la regla.

**Interesabilidad de la regla.** Mide el interés objetivo y subjetivo de la regla. Nosotros hemos utilizado la medida de interés (Tan and Kumar, 2000). En nuestra propuesta la función de evaluación tiene la siguiente expresión:

**Figura 1.2:** Medida de interés.

$$fitness(c) = \frac{w_1 * Completitud(c) + w_2 * Confianza(c) + w_3 * Interes(c)}{w_1 + w_2 + w_3}$$

Donde W1, W2 y W3 son pesos creados para el usuario, ya que el mismo puede darle mayor peso a Completitud que a Confianza e Interés. Finalmente todos los componentes de ecuación son normalizados.

El objetivo global de la función de evaluación es orientar la búsqueda hacia reglas que maximicen la precisión, minimizando el número de ejemplos negativos y no cubiertos.

### **Operadores Genéticos.**

El AG utiliza un modelo de reproducción de estado estacionario. La recombinación se realiza a través del operador de cruce multipunto y un operador de mutación uniforme sesgado con el que la mitad de las mutaciones realizadas tienen el efecto de eliminar la variable correspondiente, para incrementar la generalidad de las reglas.

### **Esquema de reproducción.**

El AG utiliza un modelo de reproducción de estado estacionario modificado (Casillas et al., 2001) que sigue el esquema siguiente: 1. Se genera una población intermedia mediante asignación de probabilidades basada en ordenación lineal y en el esquema de selección de muestreo estocástico universal de Baker. 2.

Se aplican los operadores de cruce y mutación a algunos individuos de esta población intermedia. El número de cromosomas a crear vendrá determinado por la probabilidad de cruce y mutación. 3. Los nuevos cromosomas creados sustituirán a los cromosomas peor adaptados de la población original.



De esta forma se sigue la filosofía de la reproducción estacionaria, ya que la población original sólo se modifica mediante la sustitución de los peores individuos por los individuos resultantes de la recombinación.

La aplicación de los operadores de cruce y mutación a un porcentaje de cromosomas de una población intermedia, obtenida mediante selección de los individuos mejor adaptados según un esquema de ordenación lineal y muestreo estocástico universal, implica la generación de más de dos cromosomas nuevos introduce más diversidad en la nueva población.

### 2.3. Experimentos y resultados.

Para poder utilizar el modelo de datos es necesario comenzar con la etapa de preprocesamiento de los datos y así tener una idea más clara de los conjunto de datos que analizaremos. Es de suponer que la preparación de los datos puede generar un conjunto de datos más pequeño que el original, y de esta manera mejorar la eficiencia de la Minería (ver Figura 2.1).

**Figura 2.1:** Muestra del log de Moodle después de la etapa de preprocesamiento.



```
14, 23/Jun/2006:11:10:09+0200, 201.979.148.252, 6a1b42bg4234qj1323s, GET/ugr.es, www.utem.cl
12, 23/Jun/2006:22:13:10+0200, 203.349.118.762, 76a1b42bg4234qj1323r, GET/etsitl.ugr.es, www.utem.cl
14, 23/Jun/2006:07:12:44+0200, 201.659.128.532, 5ahytrwbcy65qh3qji9oiy, GET/css/estilo.css, www.profesores/jmaroza
13, 23/Jun/2006:08:56:23+0200, 203.129.258.112, 3ehndye64ye6qhb1999w, GET/shop2.com, www.canalplus.es
14, 23/Jun/2006:21:23:15+0200, 201.349.118.562, 2v4u66b3tw5hdi809yt, GET/shop3.com, www.ya.com
15, 23/Jun/2006:03:31:13+0200, 206.239.818.242, 6a1b42bg4234qj1323o, GET/shop1.es, www.shop4.cpm
```

**Fuente:** Elaboración propia.

Al mismo tiempo que realiza la etapa de preprocesamiento se puede realizar la identificación de entradas o transacciones que va dejando el usuario durante la navegación. Es importante determinar las entradas de los usuarios para poder aplicar, en este caso el algoritmo genético multiobjetivo.

Después de realizado el preprocesamiento de los datos se procede aplicar el algoritmo genético. A continuación se muestra una regla obtenida y su interpretación.

**Si** /dt/?c=11670 **Entonces** <http://www.shop34.index/ls/index.php?&id=98> (Interés=0.51, Factor Certeza=0.79, Simpleza= 1) .

**Interpretación:** Esto indica que los usuarios visitan a la página /dt/?c=11670 y luego se van a la página <http://www.shop34.index/ls/index.php?&id=98> a demás podemos afirmar esto con una certeza del 79%. Esta regla nos muestra ciertos patrones de navegación del



usuario con el cual podemos hacer ciertas mejoras a nuestro curso para que el usuario pueda encontrar en estas páginas información mucho más interesante.

### **Conclusiones.**

- El Algoritmo Genético Multiobjetivo propuesto permite obtener reglas que facilitan la toma de decisiones sobre posibles modificaciones que se pueden realizar en los cursos, se han descrito las utilidades que pueden tener para la mejora del curso y se han mostrado ejemplos concretos de reglas descubiertas con el curso.
- La aplicación del algoritmo a un problema real ha determinado un conjunto sencillo de reglas con estructura también sencilla. Es un algoritmo que permite extraer conocimiento comprensible en forma de pocas reglas nítidas o difusas con pocas variables.
- En los trabajos futuros se debe mejorar la eficiencia del algoritmo, ya que es muy costoso computacionalmente y no se puede expandir totalmente, por lo que no se obtienen la cantidad de reglas óptimas deseadas. Se debe tratar de perfeccionar el método de discretización de modo que no se genere tanta cantidad de intervalos y en consecuencia el AG tarde en converger.

### **Referencias bibliográficas.**

- AGRAWAL, R. & SRIKANT, R. Fast algorithms for mining association rules. Proc. 20th int. conf. very large data bases, VLDB, 1994. 487-499.
- APONTE, F. A., HOYOS, J. G. & MONSALVE, J. A.. Minería de usabilidad aplicada a plataformas virtuales de aprendizaje. 2012.
- CASILLAS, J., CORDÓN, O., DEL JESUS, M. J. & HERRERA, F. Genetic feature selection in a fuzzy rule-based classification system learning process for high-dimensional problems. Information Sciences, 136, 135-157. 2001.
- CORSO, C. L. & ALFARO, S. L. MINERÍA DE USO WEB: PRESENTACIÓN DE CASO PRÁCTICO PARA EL MEJORAMIENTO DE LA EDUCACIÓN, BASADA EN EL USO DE PLATAFORMAS VIRTUALES. 2007.
- FONSECA, C. M. & FLEMING, P. J. Genetic Algorithms for Multiobjective Optimization: Formulation Discussion and Generalization. ICGA, Citeseer, 416423. 1993.
- FREITAS, A. A.. Understanding the crucial differences between classification and discovery of association rules: a position paper. AcM SIGKDD Explorations Newsletter, 2, 65-69. 2000
- FREITAS, A. A. Data mining and knowledge discovery with evolutionary algorithms, Springer Science & Business Media. 2013.
- GARCÍA, S. E., ROMERO, C., VENTURA, S. & DE-CASTRO-LOZANO, C. Sistema recomendador colaborativo usando minería de datos distribuida para la mejora continua de cursos e-learning. IEEE-RITA, 3, 19-30. 2008.

- HORTON, W. K. Designing web-based training: how to teach anyone anything anywhere anytime. New York. 2000.
- KLÖSGEN, W. & ZYTKOW, J. M. Handbook of data mining and knowledge discovery, Oxford University Press, Inc. 2002.
- LAVRAČ, N., FLACH, P. & ZUPAN, B. Rule evaluation measures: A unifying view, Springer. 1999.
- LIU, J. J. & KWOK, J. T.-Y. An extended genetic rule induction algorithm. Evolutionary Computation, 2000. Proceedings of the 2000 Congress on, 2000. IEEE, 458-463.
- MICHALEWICZ, Z. Genetic Algorithms+ Data Structures= Evolution Programs (3rd edit.) Springer. New York. 1996.
- PALMA, J. A. L. Propuesta de una metodología para perfeccionar la funcionalidad de plataformas virtuales de aprendizaje. Serie Científica, 6. 2013.
- ROMERO, C., VENTURA, S., DE CASTRO, C. & GARCÍA, E.. Algoritmos Evolutivos para Descubrimiento de Reglas de Predicción en la Mejora de Sistemas Educativos Adaptativos basados en Web 47-60. 2005a
- ROMERO, C., VENTURA, S., HERVÁS, C. & GONZÁLEZ, P.. Rule Discovery in web-based educational systems using Grammar-Based Genetic Programming. Data Mining VI: Data Mining, Text Mining and Their Business Applications, 205-214. 2005b
- SHORTLIFFE, E. H. & BUCHANAN, B. G. A model of inexact reasoning in medicine. Mathematical biosciences, 23, 351-379. 1975.
- SRIVASTAVA, J., COOLEY, R., DESHPANDE, M. & TAN, P.-N. Web usage mining: Discovery and applications of usage patterns from web data. ACM SIGKDD Explorations Newsletter, 1, 12-23. 2000.
- TAN, P. & KUMAR, V. Interesting measures for association patterns. Proc. KDD Workshop on Postprocessing in Machine Learning and Data Mining, Boston, USA, 2000. 1-9.

Para citar el artículo indexado.

Vázquez Y., Machín A., Velasteguí E., & Peñafiel V. (2018). Algoritmo genético para la generación de reglas de predicción en los cursos de moodle. *Revista electrónica Ciencia Digital* 2(1), 8-18. Recuperado desde:  
<http://www.cienciadigital.org/revistascienciadigital/index.php/CienciaDigital/article/view/1/1>



El artículo que se publica es de exclusiva responsabilidad de los autores y no necesariamente reflejan el pensamiento de la **Revista Ciencia Digital**.

El artículo queda en propiedad de la revista y, por tanto, su publicación parcial y/o total en otro medio tiene que ser autorizado por el director de la **Revista Ciencia Digital**.

