

ISSN 2410-3438

Volumen 3, Número 7 — Abril — Junio — 2016

Revista de Análisis Cuantitativo
y Estadístico

ECORFAN[®]

Indización



ECORFAN-Bolivia

- Latindex
- Google Scholar
- Research Gate
- REBID
- Mendeley
- Hispana
- Universia

ECORFAN-Bolivia

Directorio

Principal

RAMOS-ESCAMILLA, María. PhD

Director Regional

IGLESIAS-SUAREZ, Fernando. BsC

Director de la Revista

PERALTA-CASTRO, Enrique. MsC

Relaciones Institucionales

TREJO-RAMOS, Iván. BsC

Edición de Logística

CLAUDIO-MÉNDEZ, Paul. BsC

Diseñador de Edición

LEYVA-CASTRO, Iván. BsC

Revista de Análisis Cuantitativo y Estadístico, Volumen 3, Número 7, de Abril a Junio - 2016, es una revista editada trimestralmente por ECORFAN-Bolivia. Loa 1179, Cd. Sucre. Chuquisaca, Bolivia. WEB: www.ecorfan.org, revista@ecorfan.org. Editora en Jefe: RAMOS-ESCAMILLA, María, Co-Editor: IGLESIAS-SUAREZ, Fernando. ISSN: 2410-3438. Responsables de la última actualización de este número de la Unidad de Informática ECORFAN. ESCAMILLA-BOUCHÁN, Imelda, LUNA-SOTO, Vladimir, actualizado al 30 de Junio 2016.

Las opiniones expresadas por los autores no reflejan necesariamente las opiniones del editor de la publicación.

Queda terminantemente prohibida la reproducción total o parcial de los contenidos e imágenes de la publicación sin permiso del Instituto Nacional del Derecho de Autor.

Consejo Editorial

SALGADO-BELTRÁN, Lizbeth. PhD
Universidad de Barcelona, Spain

QUINTANILLA-CÓNDOR, Cerapio. PhD
Universidad Nacional de Huancavelica, Peru

GARCÍA-ESPINOSA, Cecilia. PhD
Universidad Península de Santa Elena, Ecuador

PEREIRA-LÓPEZ, Xesus. PhD
Universidad de Santiago de Compostela, Spain

VASQUEZ-GALÁN, Belem. PhD
El Colegio de la Frontera Norte, México

BANERJEE, Bidisha. MsC
Amity University, India

JESUS-NOVELO, Federico. PhD
Universidad Autónoma Metropolitana, México

SUYO-CRUZ, Gabriel. PhD
Universidad Nacional de San Antonio Abad del Cusco, Perú

Consejo Arbitral

BTO. PhD

Universidad Autónoma del Estado de México, México

APB. PhD

University of the Punjab, Pakistan

VSG. PhD

Posgrado de la Facultad de Economía, México

BGS. PhD

Universidad Complutense de Madrid, Spain.

MGM. PhD

Universidad Complutense de Madrid, Spain.

GCC. PhD

Universidad Autónoma Metropolitana - Azcapotzalco, México

MI. PhD

Instituto de Investigaciones Económicas – UNAM, México

SCJ. PhD

Universidad Juárez del Estado de Durango, México

Presentación

ECORFAN, es una revista de investigación que publica artículos en las áreas de: Revista de Análisis Cuantitativo y Estadístico.

En Pro de la Investigación, Docencia, y Formación de los recursos humanos comprometidos con la Ciencia. El contenido de los artículos y opiniones que aparecen en cada número son de los autores y no necesariamente la opinión de la Editora en Jefe.

En el primer número es presentado el artículo *Calibración y selección del modelo de aprendizaje no supervisado K-Medias, de una encuesta sobre factores de riesgo en el consumo de drogas entre estudiantes* por MEDINA-VELOZ, Gricelda, LUNA-ROSAS, Francisco Javier, TAVAREZ-AVENDAÑO, Juan Felipe y NARVAEZ-MURILLO, René Ulises, como segundo artículo está *Diseño de algoritmos en labview en un posicionador xyz para autoequipamiento y apoyo didáctico* por AGUILERA-HERNÁNDEZ, Martha Isabel, ORTIZ-SIMÓN, José Luis, RODRIGUEZ-MARTÍNEZ, Daniel y ROJO-VELASQUEZ, Gustavo, como tercer capítulo está *Efecto del raleo de frutos en el rendimiento de jitomate (Lycopersicon esculentum Mill). Bajo invernadero* por GAYTÁN-RUELAS, Marina, VARGAS-ESPINOZA, Everardo, RIVERA-ARREDONDO, Marisa y MORALES-FÉLIX, Verónica de Jesús, el cuarto artículo está *Estudio de métodos difusos para la agrupación en un conjunto de datos* presentado por FUENTES, Juan Jaime con adscripción en la Universidad Politécnica del Valle del Évora, en el quinto artículo está *Evaluación del desempeño docente en la LMS Chamilo, de las ingenierías ejecutivas de la Universidad Tecnológica Fidel Velázquez: acciones de profesionalización* presentado por HERNÁNDEZ-CRUZ, María Guadalupe, ROMERO-ROJAS, Ruth Marcela, ORTIZ-ARANGO, Víctor Tonatiuh y VARGAS-GUTIÉRREZ, Luis Daniel con adscripción en la Universidad Tecnológica Fidel Velázquez, en el penúltimo artículo está *Sistema de programación dinámica de trayectorias para asistencia de nodos bajo demanda en redes inalámbricas de sensores* presentado por GONZÁLEZ-SILVA, Marco Antonio, FRANCO-MORENO, Juan José, BAEZA-REGALADO, Michelle y MORENO-ZAMORA, Brenda Paola con adscripción en la Universidad Politécnica Metropolitana de Hidalgo y como último capítulo está *Un algoritmo transgénico vs algoritmos de optimización* por CORREA-VILLALÓN, Christian José, LUNA-RAMÍREZ, Enrique, SORIA-CRUZ, Jorge y CRUZ-VALENZUELA, Roberto con adscripción en el Instituto de Educación de Aguascalientes y en el Instituto Tecnológico El Llano Aguascalientes respectivamente.

Contenido

Artículo	Pág.
Calibración y selección del modelo de aprendizaje no supervisado K-Medias, de una encuesta sobre factores de riesgo en el consumo de drogas entre estudiantes MEDINA-VELOZ, Gricelda, LUNA-ROSAS, Francisco Javier, TAVAREZ-AVENDAÑO, Juan Felipe y NARVAEZ-MURILLO, René Ulises	1-9
Diseño de algoritmos en labview en un posicionador xyz para autoequipamiento y apoyo didáctico AGUILERA-HERNÁNDEZ, Martha Isabel, ORTIZ-SIMÓN, José Luis, RODRIGUEZ-MARTÍNEZ, Daniel y ROJO-VELASQUEZ, Gustavo	10-17
Efecto del raleo de frutos en el rendimiento de jitomate (<i>Lycopersicon esculentum</i> Mill). Bajo invernadero GAYTÁN-RUELAS, Marina, VARGAS-ESPINOZA, Everardo, RIVERA-ARREDONDO, Marisa y MORALES-FÉLIX, Verónica de Jesús	18-25
Estudio de métodos difusos para la agrupación en un conjunto de datos FUENTES, Juan Jaime	26-29
Evaluación del desempeño docente en la LMS Chamilo, de las ingenierías ejecutivas de la Universidad Tecnológica Fidel Velázquez: acciones de profesionalización HERNÁNDEZ-CRUZ, María Guadalupe, ROMERO-ROJAS, Ruth Marcela, ORTIZ-ARANGO, Víctor Tonatiuh y VARGAS-GUTIÉRREZ, Luis Daniel	30-44
Sistema de programación dinámica de trayectorias para asistencia de nodos bajo demanda en redes inalámbricas de sensores GONZÁLEZ-SILVA, Marco Antonio, FRANCO-MORENO, Juan José, BAEZA-REGALADO, Michelle y MORENO-ZAMORA, Brenda Paola	45-52
Un algoritmo transgénico vs algoritmos de optimización CORREA-VILLALÓN, Christian José, LUNA-RAMÍREZ, Enrique, SORIA-CRUZ, Jorge y CRUZ-VALENZUELA, Roberto	53-61

Instrucciones para Autores

Formato de Originalidad

Formato de Autorización

Calibración y selección del modelo de aprendizaje no supervisado K-Medias, de una encuesta sobre factores de riesgo en el consumo de drogas entre estudiantes

MEDINA-VELOZ, Gricelda*†, LUNA-ROSAS, Francisco Javier, TAVAREZ-AVENDAÑO, Juan Felipe y NARVAEZ-MURILLO, René Ulises

Recibido 21 Diciembre, 2015; Aceptado 08 Marzo, 2016

Resumen

En este trabajo se aplica un método de selección y calibración del modelo de aprendizaje no supervisado de minería de datos K-Medias, para generar el diseño de un modelo que arroje los mejores resultados dado un conjunto específico de datos, a través de un proceso denominado validación cruzada (“Cross-Validation”). El objetivo que se persigue con este proyecto, es aplicar un método de calibración y selección del modelo de aprendizaje no supervisado de minería de datos K-Medias y con sus objetivos específicos se busca: 1) Promover el uso del programa estadístico de licencia libre RStudio. 2) Analizar los datos de una encuesta sobre factores de riesgo y protección en el consumo de drogas entre estudiantes a nivel universitario. Y 3) Generar un modelo de minería de datos reutilizable del método de aprendizaje no supervisado K-Medias. La creación del modelo de minería de datos involucró el desarrollo de 4 pasos principales: 1. La creación de la estructura de minería, 2. La selección del algoritmo para el modelo. 3. La elección de los datos a incluir y 4. El procesamiento de los datos. La creación de un modelo de aprendizaje no supervisado de minería de datos del método K-Medias.

Minería de datos, aprendizaje no supervisado, K-medias, factores de riesgo en drogas

Abstract

This paper presents a method of selection and a model calibration of unsupervised learning data mining K-means, to generate the design of a model that brings the best results given a specific set of data, through a process called cross-validation. The general objective of this paper is to apply a calibration method and a model selection on unsupervised learning of data mining method called K-Means. And with the specific objectives: 1) To promote the use of statistical RStudio free license program. 2) Analyze the data of a survey on risk and protective factors in drug use among students at university level. 3) Build a model reusable data mining method on unsupervised learning K-Means. The creation of data mining model involved 4 main steps: 1. The creation of the mining structure, 2. The algorithm selection for the model. 3. The choice of the data to be included and 4. The data processing. Creating an unsupervised data mining learning model of K-means method.

Data mining, unsupervised learning, K-means, drugs usage

Citación: MEDINA-VELOZ, Gricelda, LUNA-ROSAS, Francisco Javier, TAVAREZ-AVENDAÑO, Juan Felipe y NARVAEZ-MURILLO, René Ulises. Calibración y selección del modelo de aprendizaje no supervisado K-Medias, de una encuesta sobre factores de riesgo en el consumo de drogas entre estudiantes. Revista de Análisis Cuantitativo y Estadístico. 2016. 3-7: 1-9

*Correspondencia al Autor (Correo electrónico: gricelda.medina@utna.edu.mx)

† Investigador contribuyendo como primer autor.

Introducción

Existen paquetes de estadística y minería de datos que son una especie de cajas negras donde los parámetros y los algoritmos de los métodos, no son posible manipularlos para probar sus resultados con diferentes valores.

Sin embargo, el programa estadístico de R, permite que los parámetros de algunos métodos de minería de datos, se puedan programar desde la consola y manipular sus valores y algoritmos, para así, asegurarse de obtener los mejores resultados del modelo mediante este tipo de calibración.

La finalidad del proceso de calibración de modelos de minería de datos, es buscar “maximizar la Inercia Inter-Clases, minimizar el error global, o maximizar el área bajo la curva ROC” [rpubs].

Todo depende del objetivo que se está buscando con el análisis de la información en el proyecto de minería. Para efectos de este trabajo y la evaluación del modelo de minería, se utilizó la información obtenida, de la aplicación de una encuesta sobre factores de riesgo y protección en el consumo de drogas entre estudiantes.

El tema fue elegido, debido a su relevancia actual, en los ambientes educativos, y a los resultados encontrados en la Encuesta Nacional de Consumo de Drogas en Estudiantes aplicada en el año 2014 y publicada en el año 2015.

Consumo de drogas entre estudiantes

El consumo del alcohol, de solventes inhalables, de productos de tabaco y de drogas ilegales representa un complejo fenómeno originado por un amplio entramado de factores de riesgo cuyo abordaje requiere de información epidemiológica veraz y actualizada.

Se publicó en el año 2015, la Encuesta Nacional de Consumo de Drogas en Estudiantes 2014, coordinada por la Comisión Nacional Contra las Adicciones, el Centro Nacional para la Prevención y Control de las Adicciones y el Instituto Nacional de Psiquiatría “Ramón de la Fuente Muñiz”.

En el que se encontraron resultados alarmantes del consumo de estas sustancias en estudiantes de primaria, secundaria y preparatoria. Sus datos ofrecen un amplio panorama de la situación actual y las variables asociadas, como la oportunidad de exposición, las edades de inicio y las prevalencias de consumo.

La información se obtuvo mediante un cuestionario estandarizado, y las secciones que cubría eran datos sociodemográficos, datos sobre el consumo de drogas, la conducta antisocial, el ámbito social, y el ámbito interpersonal.

Y aunque la problemática no es igual en todos los estados, ni en todos los niveles educativos o edades, sirvió para analizar los resultados obtenidos específicamente en el estado de Aguascalientes. Con los cuales se logró conocer que la prevalencia del consumo de drogas, de al menos una vez en estudiantes de secundaria es del 12.3% y en bachillerato es del 22.7%.

De los alumnos encuestados de secundaria y bachillerato, 8192 de ellos, tienen apoyo o tratamiento debido al consumo de drogas y que el 43.7% inició el consumo de drogas entre los 13 y 14 años de edad [Comisión Nacional Contra las Adicciones], por lo que se optó, por analizar a nivel universitario los factores de riesgo y protección a los cuales se enfrentan los estudiantes aún a este nivel académico.

Modelos de Minería de Datos

El propósito general de crear modelos de minería de datos, es para buscar con ello la descripción o la predicción de la información analizada [Pang-Ning], [Hand]. Con los modelos descriptivos se particionan o segmentan un conjunto de datos en grupos o conglomerados, basándose en la similaridad de ciertas variables de la información que se analiza, descubriendo de forma autónoma, correlaciones y categorías similares entre ellos, buscando que, los integrantes de cada grupos o conglomerado que se forma sean lo más acoplados o parecidos entre sí, y que los grupos sean lo más separados o diferentes posible entre ellos.

En cambio los modelos predictivos o de aprendizaje supervisado, pretenden predecir valores futuros, o desconocidos, de las variables, y su objetivo es crear una función capaz de predecir el valor correspondiente a una variable, después de haber analizado una serie de ejemplos [Hand], [Han].

Un modelo de minería de datos, se crea mediante la aplicación de un algoritmo a los datos, que después se podrá aplicar a nuevos proyectos de minería para crear predicciones y deducir relaciones y/o patrones entre ellos.

El modelo de minería, recibe los datos de una estructura de minería, los analiza utilizando el algoritmo más adecuado, y que ofrezca los mejores resultados para ese conjunto de información, para luego almacenar los resultados derivados de su procesamiento estadístico, como son los patrones encontrados y el resultado del análisis [López], [Fayad]. Para crear un modelo de minería de datos, se siguen los pasos generales descritos en la Tabla 1.

Paso	Actividad
1	Se Crea la estructura de minería de datos
2	Se selecciona el algoritmo más adecuado para la tarea analítica.
3	Se eligen las columnas que se incluirán en el modelo y se especifica su uso
4	Se rellena el modelo con datos procesando la estructura y el modelo.

Tabla 1 Pasos para crear un modelo de minería de datos

La metodología utilizada para la creación de este modelo descriptivo de minería de datos, se describe a continuación.

Metodología

1) Creación de la estructura de minería de datos: El primer paso a desarrollar en este trabajo, fue el análisis de los parámetros del método descriptivo de Clustering K-Medias, en el programa estadístico de R a través de la interfaz gráfica del programa de RStudio [Gómez], [Artime], [Chambers], [Venables], [Hothorn], con el objetivo de analizar los valores del modelo en R, y poder medir los grados de libertad o número de parámetros que se manejan en el algoritmo, los cuales se podrán calibrar y así, de esa manera crear la estructura del modelo de minería. Como se puede observar en la Figura 1, uno de los parámetros más importantes del modelo es el algoritmo a usar en el método, donde R propone el uso de 4 posibles algoritmos ("Hartigan-Wong", "Lloyd", "Forgy", "MacQueen"), para su calibración.

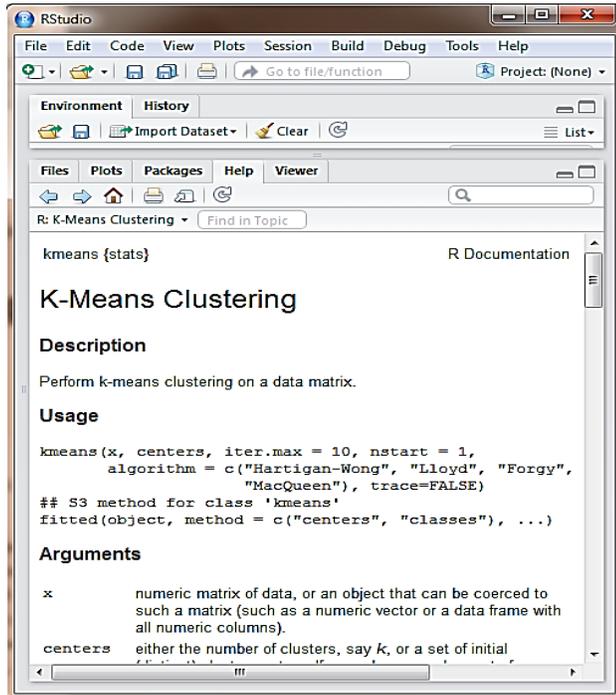


Figura 1 Estructura del modelo de minería de datos k-medias en el programa RStudio

La descripción general de cada uno de estos algoritmos se describe a continuación.

A. Método Lloyd.

Es un algoritmo propuesto por Lloyd en 1957, toma un conjunto de observaciones o casos y los clusteriza en grupos tratando de minimizar la distancia inter clusters mediante la suma de sus cuadrados [Lloyd].

$$\sum_{i=1}^k \sum_{x_j \in S_i} (x_j - \mu_i)^2 \quad (1)$$

B. Método Forgy.

Forgy, en 1965, sugiere un algoritmo simple consistente en la siguiente secuencia de pasos: 1) Comenzar con una configuración inicial. Ir al paso 2 si se comienza con un conjunto de puntos semilla.

Ir al paso 3, si se comienza con una partición de los casos. 2) Colocar cada individuo en el cluster con la semilla más próxima. Las semillas permanecen fijas para cada ciclo completo que recorra el conjunto de datos. 3) Calcular los nuevos puntos semilla como los centroides de los clusters. 4) Alternar los pasos, segundo y tercero hasta que el proceso converja, es decir, continuar hasta que ningún individuo cambie de cluster en el paso segundo [Forgy].

C. Método MacQueen.

MacQueen, en 1972, emplea el término K-Medias para denotar el proceso de asignar cada individuo al cluster (de los K prefijados), con el centroide más próximo. La clave de este procedimiento radica en que el centroide se calcula a partir de los miembros del cluster, tras cada asignación y no al final de cada ciclo, como ocurre en los métodos de Forgy y Jancey.

El algoritmo de MacQueen considera la siguiente secuencia de pasos: 1. Toma los K primeros casos como clusters unitarios. Luego 2. Asigna cada uno de los $m - K$ individuos restantes al cluster con el centroide más próximo, y después de cada asignación, recalcula el centroide del cluster obtenido.

Finalmente, 3. Tras la asignación de todos los individuos en el paso segundo, toma los centroides de los clusters existentes como puntos semillas fijos, y hace una pasada más sobre los datos asignando cada dato al punto semilla más cercano.

El último paso, es el mismo que el del método de Forgy, excepto que la recolocación se efectúa una vez más sin esperar a que se produzca la convergencia [MacQueen].

D. Método Hartigan and Wong.

El objetivo del algoritmo K-medias, es dividir M puntos en N dimensiones dentro de K conglomerados, de modo que se minimiza la suma dentro de los cuadrados. Pero esto no es práctico, excepto cuando M y N son pequeños y $K = 2$. En cambio, con el algoritmo de Hartigan se busca un óptimo local, que solucione que no se muevan de un punto de un grupo a otro, reduciendo la suma inter-clusters de los cuadrados [Hartigan].

$$Sum(k) = \sum_{i=0}^n \sum_{j=0}^p (x(i, j) - x(k, j))^2 \quad (2)$$

Considerando entonces, que el método k-medias en el programa estadístico de R, puede ser calibrado con cada uno de estos algoritmos. La estructura del modelo de minería se construyó, probando el método con cada uno de ellos, para comparar sus resultados y elegir aquel que arrojó mejores resultados para este conjunto específico de datos.

Factor	Elemento de riesgo o protección
1	Malestar emocional
2	Satisfacción en las relaciones personales
3	Concepto y valoración de las drogas
4	Espiritualidad
5	Permiso social y acceso a las drogas
6	Habilidad social y auto control

Tabla 2 Factores de riesgo y protección en el consumo de drogas

Los datos utilizados para hacer las pruebas del modelo de minería, se obtuvieron de la aplicación de un cuestionario a los estudiantes, y éste se estructuró considerando como base el trabajo propuesto por un grupo de sicólogos colombianos sobre la elaboración de un sondeo de factores de riesgo y de protección para el consumo de drogas en jóvenes universitarios, publicado en el 2006, por la revista de la Universidad Católica de Colombia.

El cuestionario se aplicó a 60 alumnos de la carrera de Tecnologías de la Información y la Comunicación, de la Universidad Tecnológica del Norte de Aguascalientes. Los factores que se analizaron con el cuestionario, se muestran en la Tabla 2.

2) Selección del algoritmo: En la selección del algoritmo más adecuado, y que ofrece los mejores resultados para este conjunto de datos, se utilizaron las técnicas estadísticas de cross validation, técnicas propuestas por los doctores Bradley Efron y Rob Tibshirani de la Universidad de Standford [Efron], [Friedman]. Ellos proponen que, para obtener una adecuada calibración del modelo, primeramente se debe iniciar con la selección del número de clusters (k = número de grupos que se crearán). Este proceso es automático, cuando se utiliza en minería de datos el método de clasificación jerárquica, debido a que, el árbol binario que se genera, sugiere automáticamente cuantos clusters se deben de elegir. Pero para el caso de k-medias, existe un cálculo llamado el codo de Jambú, que es el que ayuda a determinar, el número de clusters que se deben elegir para el método [Jambú], [Tibshirani]. El problema con esto es, que su cálculo requiere de muchas operaciones, ya que implica ir graficando la inercia intra clases, del proceso de calibración del modelo. El nombre de codo, lo recibe en base a la forma de la gráfica que se genera y, el punto donde comienza su estabilización o disminuye la variación, indica donde se encuentra el k ideal, que hace que la inercia se estabilice, lo que implica, que ya no sea necesario tratar de mejorar porque la inercia intra clases se ha estabilizado.

```

Calibracion_K_Medias.R x
Source on Save Run Source
1 setwd("C:/Gris/MD")
2 datos <- read.csv("Datos.csv",header=TRUE, sep=";", dec=".", row.names=1)
3 dim(datos)
4 InerciaIC.Hartigan=rep(0,30)
5 InerciaIC.Lloyd=rep(0,30)
6 InerciaIC.Forgy=rep(0,30)
7 InerciaIC.MacQueen=rep(0,30)
8 For(k in 1:30) {
9   grupos=kmeans(datos,k,iter.max=100,algorithm = "Hartigan-wong")
10  InerciaIC.Hartigan[k]=grupos$tot.withinss
11  grupos=kmeans(datos,k,iter.max=100,algorithm = "Lloyd")
12  InerciaIC.Lloyd[k]=grupos$tot.withinss
13  grupos=kmeans(datos,k,iter.max=100,algorithm = "Forgy")
14  InerciaIC.Forgy[k]=grupos$tot.withinss
15  grupos=kmeans(datos,k,iter.max=100,algorithm = "MacQueen")
16  InerciaIC.MacQueen[k]=grupos$tot.withinss
17 }
18
19 plot(InerciaIC.Hartigan,col="blue",type="b")
20 points(InerciaIC.Lloyd,col="red",type="b")
21 points(InerciaIC.Forgy,col="green",type="b")
22 points(InerciaIC.MacQueen,col="magenta",type="b")
23 legend("topright",legend = c("Hartigan","Lloyd","Forgy","MacQueen"),
24       col = c("blue","red","green","magenta"), lty = 1, lwd = 1)
25:1
R Script

```

Figura 2 Calibración del modelo k-medias en el programa RStudio [rpubs]

Para generar el código de Jambú de este proyecto, se corrieron 30 veces el algoritmo k-medias, como una forma de asegurarse que el resultado obtenido sea el más adecuado, recordando que si se corre el proceso de k-medias una vez y luego se corre una vez más, el resultado puede ser considerablemente diferente, ya que el método arranca con datos aleatorios elegidos automáticamente, Gráfico 1.

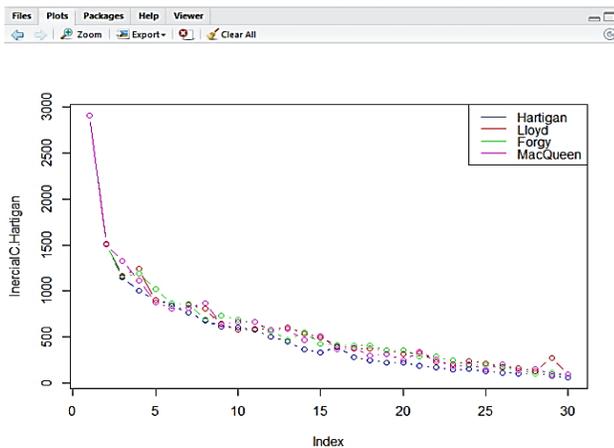


Gráfico 1 Codo de Jambú

En el proceso se creó un ciclo que se ejecutó 30 veces, y en cada corrida del ciclo, el método k-medias se programó con un algoritmo diferente ("Hartigan-Wong", "Lloyd", "Forgy", "MacQueen").

El resultado de la inercia se guardó en su respectivo vector y posición, por medio de la variable `grupos$tot.withinss` que almacena la inercia intra clases Figura 2.

```

Console C:/Gris/MD/
> Hartigan<-0
> Lloyd<-0
> Forgy<-0
> MacQueen<-0
> for(i in 1:50) {
+   grupos<-kmeans(datos,5,iter.max=100,algorithm = "Hartigan-wong")
+   Hartigan<-Hartigan+grupos$betweenss
+   grupos<-kmeans(datos,5,iter.max=100,algorithm = "Lloyd")
+   Lloyd<-Lloyd+grupos$betweenss
+   grupos<-kmeans(datos,5,iter.max=100,algorithm = "Forgy")
+   Forgy<-Forgy+grupos$betweenss
+   grupos<-kmeans(datos,5,iter.max=100,algorithm = "MacQueen")
+   MacQueen<-MacQueen+grupos$betweenss
+ }
> Hartigan/50
[1] 2006.719
> Lloyd/50
[1] 1963.549
> Forgy/50
[1] 1964.973
> MacQueen/50
[1] 1969.767
> |

```

Figura 3 Selección del algoritmo en el modelo k-medias del programa RStudio [rpubs]

Después se plotearon los vectores de las inercias de cada algoritmo para producir el gráfico que se muestra en la Gráfico 1, y que representa el código de Jambú, con el cual se buscó, el punto de estabilización para encontrar el mejor valor de k, y encontrar el número de clusters que se recomienda utilizar en el modelo de minería de datos para este juego específico de datos. Después de seleccionar, mediante la ayuda del código de Jambú, el número de clusters que generaría el modelo, se hizo una nueva calibración, ahora para elegir cuál de los algoritmos produce los mejores resultados. Para ello se definieron variables que sirvieron como acumuladores de inercias intra clases en cada uno de los algoritmos del modelo. Luego se creó el algoritmo que se corrió 50 veces y en cada corrida se almacenó en el acumulador las inercias intra clases. Finalmente se promedió cada acumulador entre el número de iteraciones que corrió el ciclo. Para luego elegir el algoritmo que dio una inercia inter clases mayor. Como puede observarse en la Figura 3 el algoritmo que arrojó los mejores resultados para este juego específico de datos es el algoritmo de Hartigan y Won.

	A	B	C	D	E	F	G
1	Alu	MalestaEmocional	SatisRelPersonales	ConceptoYValorDeDrogas	Espiritualidad	AccesoADrogas	PermisoSocialYAutoControl
2	a1	3	5	2	4	3	3
3	a2	2	4	2	3	3	4
4	a3	0	5	2	5	1	4
5	a4	2	4	1	5	3	3
6	a5	1	3	3	1	1	3
7	a6	1	4	2	4	3	3
8	a7	2	3	3	3	3	2
9	a8	3	2	2	5	2	2

Figura 4 Estructura del archivo .csv, resultado de la encuesta aplicada sobre factores de riesgo y protección en el consumo de drogas

3) Elección de las columnas a incluir en el modelo. Para la elección del número de columnas a incluir en el modelo, se determinó que, todas las columnas son necesarias para la evaluación y construcción del mismo, ya que el archivo de datos que se usó para probar el modelo, es un archivo .csv, con solo 7 columnas, donde la columna 1 representa los alumnos a los cuales les fue aplicada la encuesta, y las columnas 2-7 muestran cada uno de los factores que se analizaron del tema objeto de estudio, los cuales se mencionaron en la Tabla 2 del documento. La estructura de dicho archivo, se muestra en la Figura 4.

4) Llenado del modelo y procesamiento de la estructura. El archivo del modelo se almacenó con el nombre de Calibracion_K_Medias.R, para su uso posterior en la evaluación de nuevos análisis descriptivos de datos.

Con los que solo será necesario cargar el nuevo archivo .csv con los nuevos datos y el modelo generará automáticamente los resultados que arrojan cada uno de los algoritmos de calibración del método k-medias. Los cuales indicarán, cuál de los cuatro algoritmos se recomienda usar, para la construcción del modelo de minería, que ofrezca los mejores resultados para ese nuevo juego específico de datos.

Resultados obtenidos

En base a los resultados obtenidos en el proceso de calibración del modelo de minería, y como se puede observar en la Figura 3, el algoritmo con el cual se obtuvieron mejores resultados, fue el algoritmo propuesto por Hartigan, ya que este algoritmo reportó una mayor inercia inter clases, lo que significa que los grupos o clusters formados con este algoritmo tienen más semejanzas y similitudes entre sus integrantes, y que los grupos entre ellos muestran mayores diferencias que los grupos generados por los otros algoritmos. El Gráfico 2, muestra la gráfica de araña que generó el modelo minería, y en la que se puede observar, las características que identifican a cada cluster creado.

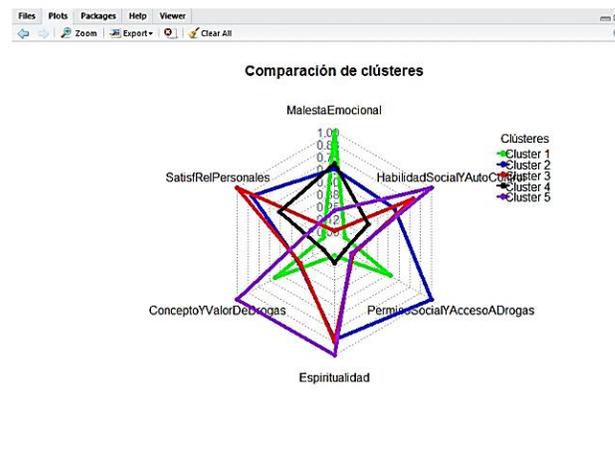


Gráfico 2 Comparativa de clusters en forma de araña, resultado del modelo k-medias

En base a esta gráfica se construyó la Tabla 3 de resumen, donde se indican en la columna 3 las características que identifican a cada cluster, en base al concepto analizado en la columna 2. Por ejemplo, el factor de malestar emocional está más presente en los clusters 1, 2 y 4.

Factor	Concepto	Cluster
1	Malestar emocional	1,2,4
2	Satisfacción en las relaciones personales	3,2,4
3	Concepto y valoración de las drogas	5,1
4	Espiritualidad	5,2,3
5	Permiso social y acceso a las drogas	1,2
6	Habilidad social y auto control	5,3,4

Tabla 3 Identificación de clusters en base a los factores de riesgo y protección en el consumo de drogas

Los clusters que piensan que las drogas son permitidas socialmente y tienen un mayor acceso a ellas, son los integrantes de los clusters 1 y 2. Y así sucesivamente se pueden analizar y valorar las características de cada cluster formado por el modelo de minería. Esto nos puede dar una referencia para la elaboración de programas y/o acciones a considerar, los cuales se pueden preparar para ayudar y disminuir el alto grado consumo de drogas entre los estudiantes, a cualquier nivel académico.

Otra de las gráficas que arroja el modelo de minería, es el plot de características de clusters, Gráfico 3, el cual puede ser útil para hacer comparaciones del total de clusters con mayor facilidad. Por ejemplo se puede observar fácilmente, que el cluster 3, es el cluster con menor nivel de malestar emocional contrario a lo que reporta el cluster 1, con el mayor nivel en este aspecto. El cluster 2 es el que reporta más permiso social y acceso a drogas, y el cluster 5 es el que reporta un mayor nivel de espiritualidad, habilidad social y auto control, etc.

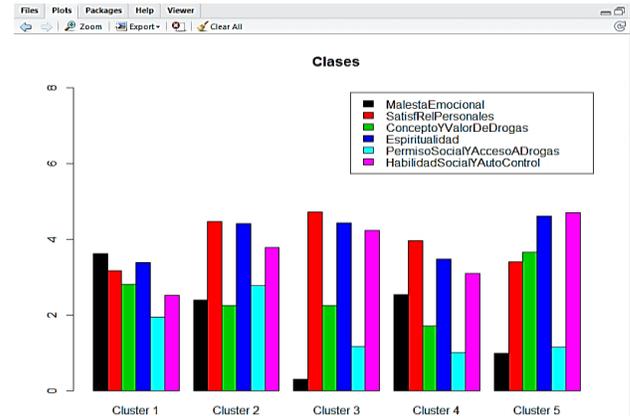


Gráfico 3 Plot comparativo de clusters, resultado del modelo k-medias

Conclusiones

En este trabajo se mostró, cómo con el uso de la técnica de validación cruzada (cross validation) se puede calibrar un método exploratorio de minería de datos, (este caso el método k-medias), para construir y seleccionar el mejor modelo de minería que arroje los mejores resultados, para un juego específico de datos. Además de que, el modelo continúa siendo útil para la evaluación de nuevos datos de minería que utilicen el método de k-medias. Solo faltaría la preparación adecuada de los datos, cargar los datos en el modelo y este automáticamente generará los resultados, calibrados con cada uno de los algoritmos posibles con los que se puede trabajar el método en el programa de RStudio. Con lo cual se asegura que el modelo está construido con los mejores parámetros que el método puede ofrecer y por ende, generar mejores resultados con un mayor índice de confiabilidad.

Referencias

- [Arttime] Arttime, C. C., & Blanco, N. C. (2013). Paquetes estadísticos con licencia libre (I)/Free statistical software (I). REMA, 18(2), 12-33.
- [Chambers] Chambers, J. (2008). Software for data analysis: programming with R. Springer Science & Business Media.

[Comisión Nacional Contra las Adicciones] Instituto Nacional de Psiquiatría Ramón de la Fuente Muñiz; Comisión Nacional Contra las Adicciones, Secretaría de Salud. Encuesta Nacional de Consumo de Drogas en Estudiantes 2014: Reporte de Drogas. Villatoro-Velázquez JA, Oliva Robles, N., Fregoso Ito, D., Bustos Gamiño, M., Mujica Salazar, A., Martín del Campo Sánchez, R., Nanni Alvarado, R. y Medina-Mora ME. México DF, México: INPRFM; 2015.

[Efron] Efron, B., & Gong, G. (1983). A leisurely look at the bootstrap, the jackknife, and cross-validation. *The American Statistician*, 37(1), 36-48.

[Fayyad] Fayyad, U., Piatetsky-Shapiro, G., & Smyth, P. (1996). From data mining to knowledge discovery in databases. *AI magazine*, 17(3), 37.

[Forgy] Forgy, E. W. (1965). Cluster analysis of multivariate data: efficiency versus interpretability of classifications. *Biometrics*, 21, 768-769.

[Friedman] Friedman, J., Hastie, T., & Tibshirani, R. (2001). *The elements of statistical learning* (Vol. 1). Springer, Berlin: Springer series in statistics.

[Gómez] Gómez, A. A. (2008). *Estadística básica con R y R-Commander*. Servicio Publicaciones UCA.

[Han] Han, J., Pei, J., & Kamber, M. (2011). *Data mining: concepts and techniques*. Elsevier.

[Hothorn] Hothorn, T., & Everitt, B. S. (2014). *A handbook of statistical analyses using R*. CRC press.

[Hand] Hand, D. J., Mannila, H., & Smyth, P. (2001). *Principles of data mining*. MIT press.

[Hartigan] Hartigan, J. A., & Wong, M. A. (1979). Algorithm AS 136: A k-means clustering algorithm. *Journal of the Royal Statistical Society. Series C (Applied Statistics)*, 28(1), 100-108.

[Jambu] Jambu, M., Tan, S. H., & Stern, D. N. (1989). *Exploration informatique et statistique des données*. Dunod.

[Jambu] Jambu, M. (1991). *Exploratory and multivariate data analysis*. Elsevier.

[López] López, C. P. (2007). *Minería de datos: técnicas y herramientas*. Editorial Paraninfo.

[Lloyd] Lloyd, S. (1982). Least squares quantization in PCM. *IEEE transactions on information theory*, 28(2), 129-137.

[MacQueen] MacQueen, J. (1967, June). Some methods for classification and analysis of multivariate observations. In *Proceedings of the fifth Berkeley symposium on mathematical statistics and probability* (Vol. 1, No. 14, pp. 281-297).

[Pang-Ning] Pang-Ning, T., Steinbach, M., & Kumar, V. (2006). *Introduction to data mining*. In *Library of congress* (Vol. 74).

[rpubs] <http://rpubs.com/orodriguez/13318>

[Tibshirani] Tibshirani, R., Walther, G., & Hastie, T. (2001). Estimating the number of clusters in a data set via the gap statistic. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 63(2), 411-423.

[Venables] Venables, W. N., & Ripley, B. P. (1990). *R: A Programming Environment for Data Analysis and Graphics*. Technical Report, Department of Statistics, University of Adelaide (Dec.).

Diseño de algoritmos en labview en un posicionador xyz para autoequipamiento y apoyo didáctico

AGUILERA-HERNÁNDEZ, Martha Isabel*†, ORTIZ-SIMÓN, José Luis, RODRIGUEZ-MARTÍNEZ, Daniel y ROJO-VELASQUEZ, Gustavo

Recibido Octubre 06, 2015; Aceptado Febrero 18, 2016

Resumen

Esta investigación analiza con un diseño cuantitativo la diferencia en el estilo de liderazgo percibido y preferido por los seguidores de tres cohortes generacionales trabajando en la industria maquiladora en México. Los baby boomers, la generación X y la generación Y, confluyen en un ambiente laboral en que de sus interrelación dependerá el éxito de las actuales y futuras empresas. Siguiendo el diseño de Cohen, se realizaron 68 encuestas por cohorte generacional. La hipótesis asociada con la pregunta de investigación fue probada mediante la prueba de los rangos con signo de Wilcoxon. La normalidad de los datos fue tratada con la prueba estadística de Kolmogorov-Smirnov (K-S). Los resultados de este estudio encontraron diferencias entre el liderazgo percibido y preferido en el cohorte generacional de los baby boomers. Por lo que las empresas requieren estar atentas a estas diferencias generacionales. Los resultados de esta investigación demuestran y confirman la naturaleza integral e interdependiente de los conceptos clave que se utilizan como fundamentos teóricos para el estudio del entorno operativo en el desarrollo del liderazgo. Las implicaciones prácticas de los resultados son sujetas a discusión. Las oportunidades para futuras investigaciones y las limitaciones del estudio son presentadas.

Recursos humanos, liderazgo, cohortes generacionales

Abstract

This research analyzes with a quantitative design, the difference in the leadership style perceived and preferred by followers of three generational cohorts working in the maquiladora industry in Mexico. The baby boomers, generation X and generation Y, converge in a work environment in which their interaction depend on the success of current and future businesses. Following the design of Cohen, 68 surveys were conducted by generational cohort. The question associated with the research hypothesis was tested by the Wilcoxon signed-rank. The normality of the data was treated with the Kolmogorov-Smirnov (K-S) statistical test. The results of this study found differences between perceived and preferred leadership by followers of the generational cohort of baby boomers. So companies need to be aware of these generational differences. The results of this research demonstrate and confirm the integral and interdependent nature of the key concepts used as a theoretical basis for the study of the operating environment in leadership development. The practical implications of the results are subject to discussion. Opportunities for future research and the limitations of the study are presented.

Human Resources, leadership, generational cohorts

Citación: AGUILERA-HERNÁNDEZ, Martha Isabel, ORTIZ-SIMÓN, José Luis, RODRIGUEZ-MARTÍNEZ, Daniel y ROJO-VELASQUEZ, Gustavo. Diseño de algoritmos en labview en un posicionador xyz para autoequipamiento y apoyo didáctico. Revista de Análisis Cuantitativos y Estadísticos. 2016. 3-7: 10-17

*Correspondencia al Autor (Correo Electrónico: aguilera@itnuevolaredo.edu.mx)

† Investigador contribuyendo como primer autor

Introducción

Las herramientas de software para el control de sistemas de posicionamiento ha presentado una creciente demanda debido a la necesidad de obtener alta eficiencia. Las principales ventajas que presentan estos sistemas son: la reducción de costos y el incremento de precisión, lo que da como consecuencia una mayor calidad y productividad. En la actualidad se tienen sistemas mecánicos de posicionamiento que requieren ser automatizados con la finalidad de reducir los costos de adquisición de uno nuevo. Gran parte de estos equipos no están obsoletos en sus partes mecánicas pero sí en la parte de control.

Las redes de comunicación actuales requieren de programas de software compatibles para poder enviar y controlar los dispositivos mecánicos del sistema automatizado. Nuestra institución cuenta con un sistema posicionador en tres dimensiones que fue diseñado como modelo industrial en 1990. Con la finalidad de implementar nuevas tecnologías en el aprendizaje y las diferentes áreas de desarrollo de la ingeniería, se diseñó para este sistema una librería de algoritmos de posicionamiento en Labview.

Esto le permitirá al equipo actualizarse para que pueda ser utilizado como célula en una línea de producción. Toda esta actualización tiene un objetivo didáctico como apoyo para la realización de prácticas de aplicaciones industriales y a la vez será un ejemplo de autoequipamiento en nuestra institución.

El artículo presenta primeramente los antecedentes, después el desarrollo mostrando en este rubro las consideraciones técnicas del equipo. Después se muestran los resultados y las conclusiones del trabajo realizado.

Antecedentes

El posicionador en tres dimensiones fue desarrollado utilizando motores de pasos, lo que permite exactitud en el posicionamiento. Es aplicable en cualquier tipo de máquina que requiera posicionamiento como ejemplo: una fresadora. En este caso el eje z le permitiría al sistema controlar la profundidad que se requiere de la herramienta en el material. El sistema de control diseñado en 1990 se basó en un micro 8085, una actualización se realizó en el 2014 con la integración de un microcontrolador Arduino, para integrarlo en proyectos multidisciplinarios. Actualmente se ha estado trabajando en la integración de una base de datos de programas en Labview, con la finalidad de que se convierta en apoyo didáctico en las materias de relacionadas a automatización.



Figura 1 Equipo posicionador en los ejes XYZ

Desarrollo del diseño de algoritmos

El diagrama descriptivo del sistema posicionador se muestra en la figura 2. El sistema contiene el microcontrolador arduino y tres motores a pasos uno para cada uno de sus ejes tridimensionales (x, y, z). Donde cada vuelta que realice son 0.38cm linealmente. La computadora contiene la base de programas en Labview para el control de los motores.

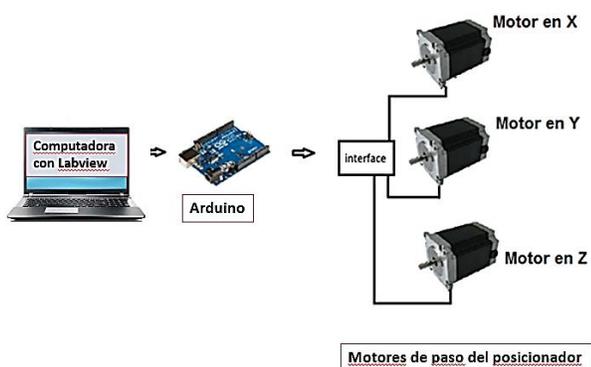


Figura 2 Componentes del sistema posicionador

Consideraciones técnicas: El motor del eje Z, el cual se muestra en la figura 3, tiene un mecanismo de baleros para disminuir la fricción y mejorar el desplazamiento (figura 4).



Figura 3 Motor y banda ensamblados



Figura 4 Sistema de baleros para el eje Z

Para realizar el control se utiliza el arduino UNO. En total se necesitan 12 salidas de la tarjeta Arduino, dado que cada motor requiere 4 entradas para señal. La figura 5, muestra las salidas utilizadas en el arduino para controlar los motores. El motor X, a los pines 2,3,4,5. El motor Y a los pines 6,7,8,9. El motor Z, a los pines 10,11,12,13.



Figura 5 Conexiones de la tarjeta Arduino

Para la comunicación entre el microcontrolador Arduino y los motores de pasos se tiene una etapa de potencia debido al alto consumo de corriente de estos. Dicha etapa contiene dos tipos de transistores: El 2N3055, que es un transistor NPN de potencia diseñado para aplicaciones de propósito general y el 2N2222, que es un transistor bipolar NPN de baja potencia de uso general, sirve tanto para aplicaciones de amplificación como de conmutación.

En la figura 6 se muestra el diagrama esquemático del circuito para un motor. El pulso entra por la base del transistor 2N2222A y este manda la señal hacia a la base del transistor de alta potencia 2N3055, que conmuta para accionar la bobina del motor, cada vez que se genere la señal de salida.

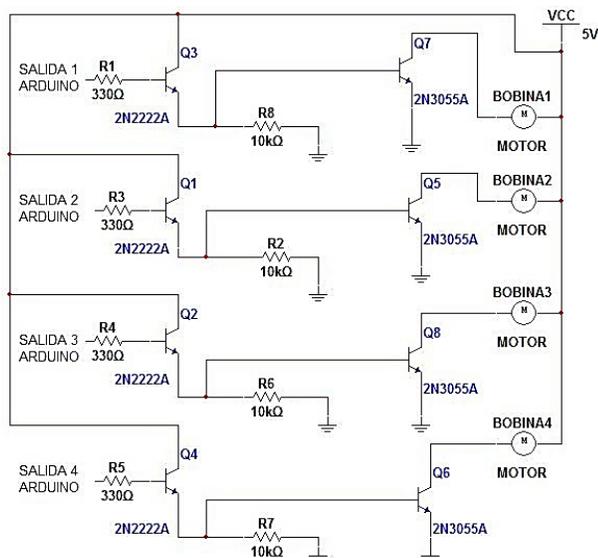


Figura 6 Diagrama esquemático para accionar un motor

Los transistores están colocados en una base disipadora (Figura 7), la cual sirve para eliminar el calor generado por la alta corriente que maneja cada uno de los motores.



Figura 7 Módulo de transistores

Diseño de algoritmos en labview

El programa LabVIEW (Laboratory Virtual Instrumentation Engineering Workbench), es una plataforma que permite diseñar algoritmos con un lenguaje de programación visual gráfico. Generalmente se utiliza como software de pruebas, de control y diseño de sistemas industriales.

En este trabajo se diseñó un conjunto de algoritmos para formar un módulo de graficación. El cual tuviera cuatro opciones de figuras geométricas: Cuadrado, rectángulo, rombo y triángulo. Y un módulo libre de archivo, en el cual se puede guardar una secuencia dada por el usuario y después reproducirse en el sistema posicionador. Esto permite que puedan guardarse diferentes maquinados de piezas, las cuales no sean precisamente las figuras geométricas definidas en el módulo de graficación. La figura 9, muestra el diagrama de flujo del módulo de graficación en el cual se requiere del usuario que introduzca las dimensiones de la figura a realizar en centímetros. La pantalla en Labview del módulo graficador se muestra en la figura 9. La figura 10, muestra el diagrama de flujo del módulo de graficación libre. En esta opción, el usuario utilizará las teclas de flecha de la computadora para el movimiento del posicionador en los ejes X ($\rightarrow\leftarrow$), Y ($\uparrow\downarrow$) y Z (RePag, AvPag). Las posiciones realizadas se guardan en un archivo, en el cual el usuario puede guardarlo para utilizarlo cuando desee. Las posiciones se repetirán de acuerdo a las posiciones realizadas por el usuario.

Diagrama de Flujo General
Del Módulo de Graficación

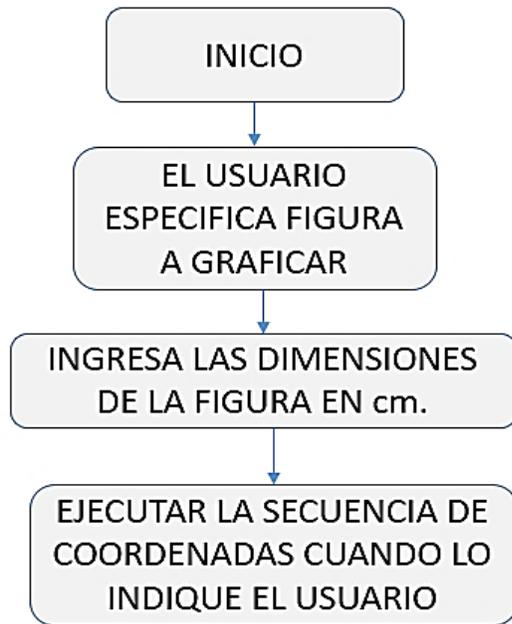


Figura 8 Diagrama de flujo general del módulo de graficación

Diagrama de Flujo General
Del Módulo Libre

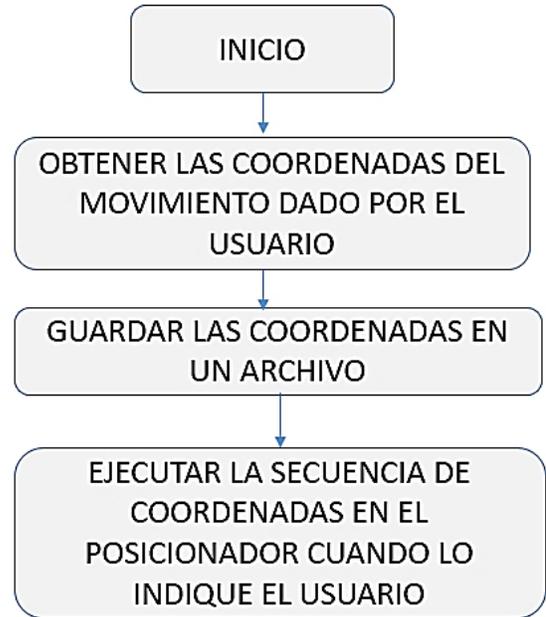


Figura 10 Diagrama de flujo general del módulo libre con archivo

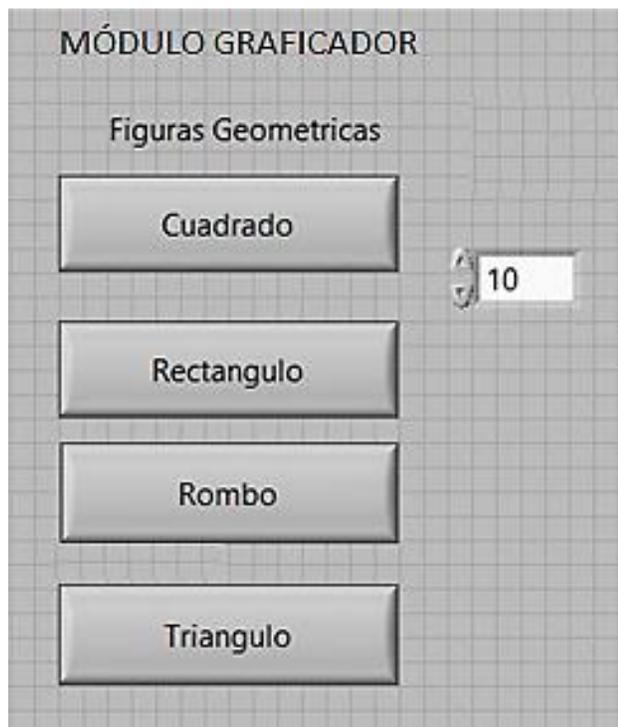


Figura 9 Tablero en Labview del módulo graficador

La comunicación entre Labview y Arduino se realizó mediante un programa denominado “LIFA_Base”, viene con un toolkit de comunicaciones.

Este programa se puede descargar directamente de la página de internet de National Instruments.

Con esta herramienta, se pueden envían y recibir datos del microcontrolador arduino, de este manera, se pueden crear las rutinas de control para los motores del posicionador. En la figura 10, se muestran algunos diagramas base que son utilizados en los algoritmos diseñados en Labview. La figura 11, muestra un programa en Labview para el control de un motor.

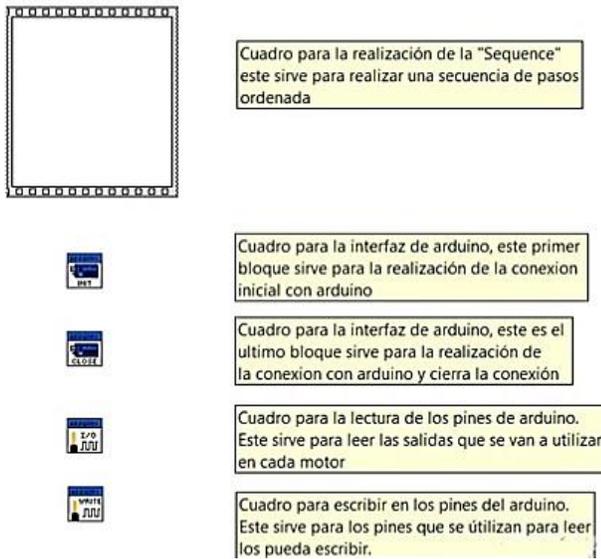


Figura 11 Diagramas base de comunicación con el arduino que fueron utilizados en los algoritmos diseñados en Labview

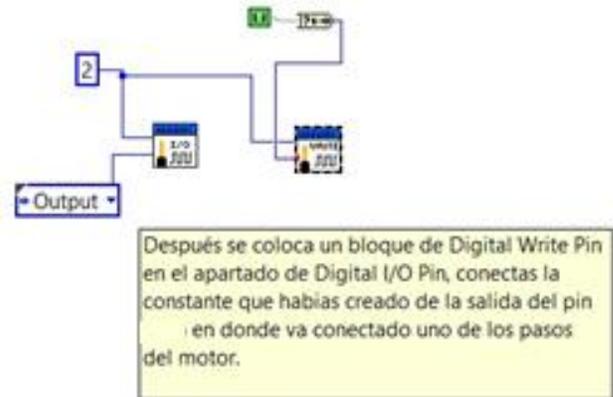


Figura 14 Utilización de la función "Digital Write Pin"

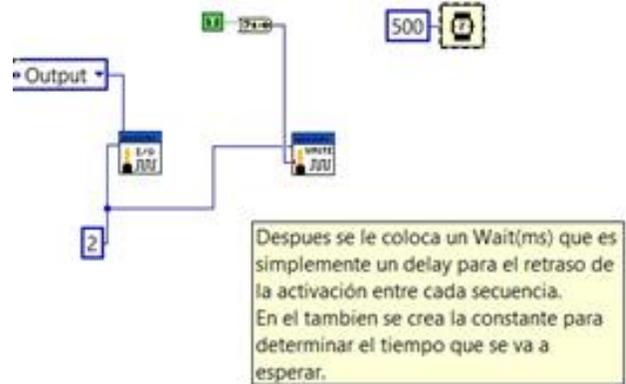


Figura 15 Utilización de la función "Wait(ms)"

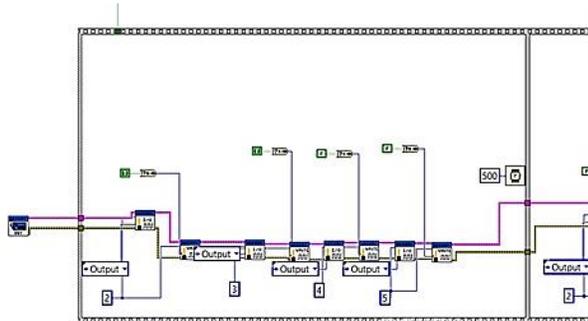


Figura 12 Ejemplo de Programa VI de Labview para el control de un motor



Figura 13 Utilización de la función "Set Digital Mode"

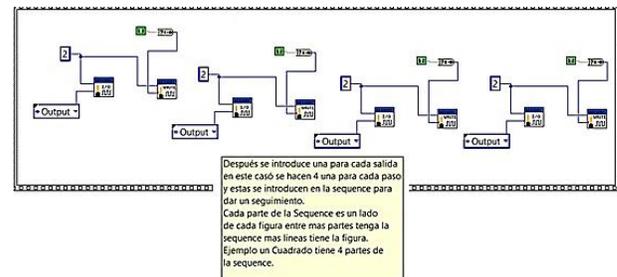


Figura 16 Utilización de la función "Sequence"

A continuación se presenta un ejemplo de diseño de algoritmo de un cuadrado.

Ejemplo de diseño de algoritmo para la figura "cuadrado" para el posicionador controlado desde Labview:

- Se posiciona en la ventana de diagrama de bloques.
- Se presiona click derecho y se abre la ventana de comandos.
- Se posiciona en “arduino low lever”.
- Se coloca en “Digital Set Pin” y click derecho en las opciones y se coloca un control. (Figura 13)
- Se introduce un “Digital Write” y se une a las opciones con las del “Digital Set Pin”. (Figura 14)
- Se coloca un set de “Digital Set Pin” y “Digital Write” para cada salida del motor en este caso son 4.
- Click derecho Boolean. “True block”, y “To(0,1)”
- Estos se unen y se anexan a “Digital Write” en la sección de Value
- Click derecho “Timer” y se coloca un waits(ms). (Figura 15)
- Click derecho “Arrays” y se coloca un “Sequence Structure”.(Figura 16)
- En la “Sequence Structure” click derecho add frame
- Coloca 16 add frame, son 4 para cada lado del cuadrado
- Asigna la salida del arduino para ambos motores. (Figura 17)
- En la secuencia booleana se coloca para cada lado (Figura 18)
- Se checa que todas las conexiones estén correctas, se corre el programa.

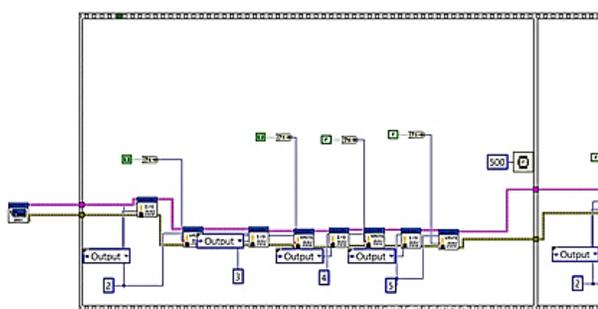


Figura 17 Diagrama del motor para el seguimiento de un cuadrado

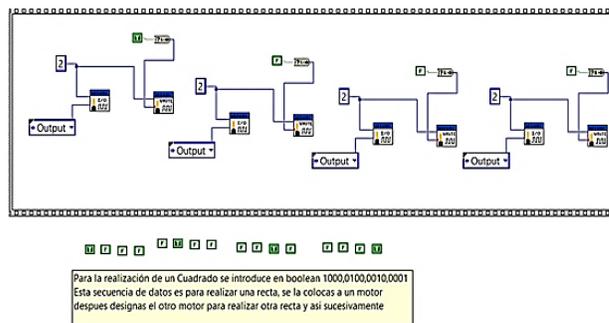


Figura 18 Pasos booleanos para la elaboración del cuadrado

Para las medida de los lados se tiene una equivalencia de 0.38cm por cada vuelta en todas las figuras. La elaboración de un triángulo se realizó con el mismo sistema “Sequence” pero esta vez en paralelo para que activara dos motores al mismo tiempo y se realizara una recta tangente al punto de origen.

Resultados

Se obtuvieron dos algoritmos: un módulo graficador en el cual se encuentran cuatro figuras geométricas y un módulo de graficación libre en el cual el usuario puede indicar los movimientos del posicionador directamente y se guardan en un archivo para poder reproducirse cuando se requiera. Estos algoritmos servirán de apoyo didáctico en las materias relacionadas a automatización y control. El posicionador realiza las tareas planteadas por el usuario y esto ha permitido que el uso que se le da al equipo sea más amplio por medio de la interfaz gráfica que permite Labview. Este equipo es un ejemplo de auto equipamiento para nuestra institución. De otra manera, el equipo se hubiera quedado sin uso ya que los programas que utilizaba anteriormente, aunque bastante eficientes en el sentido de realizar la actividad de maquinado, le faltaba la facilidad de enlace con el usuario y demás equipos de producción más actuales.

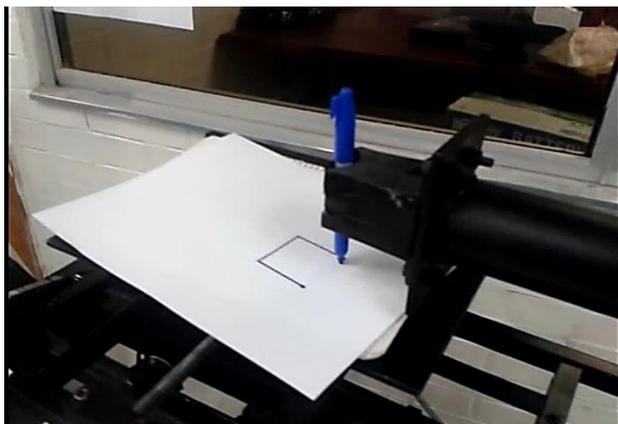


Figura 19 Posicionador realizando un cuadrado

Conclusiones y recomendaciones

Los resultados obtenidos nos permiten concluir que el diseño de algoritmos en Labview para este tipo de equipo es de suma importancia en proyectos que involucren diseño de control y de instrumentación. El proyecto en sí, es ejemplo de un programa de autoequipamiento que se tiene en el área, que incluye la actualización de equipo que funcionaba con software obsoleto. Este tipo de actualización permite la utilización del equipo mediante las herramientas de software actuales. El diseño de todas estas librerías serán aplicadas a otros equipos con las debidas modificaciones. Esto permitirá que el software pueda crecer y aplicarse en áreas de especialidad de robótica.

Referencias

Aguilera Hdz. Martha, Tesis de Maestría "Posicionador en tres dimensiones", junio de 1990.

Arduino, <http://arduino.cc/es/>, Arduino UNO, 15 de Septiembre del 2016

Evans Brian W, Arduino programming notebook, Creative Commons, Agosto 2007

Guel Ramiro, Archivo de Residencias Profesionales "Implementación de un microcontrolador en un sistema posicionador en tres dimensiones (x, y, z)", 4 de Diciembre del 2014

National Instruments, <http://www.ni.com/labview/esa/>, Labview, 15 de Septiembre del 2016

Efecto del raleo de frutos en el rendimiento de jitomate (*Lycopersicon esculentum* Mill). Bajo invernadero

GAYTÁN-RUELAS, Marina*†, VARGAS-ESPINOZA, Everardo, RIVERA-ARREDONDO, Marisa y MORALES-FÉLIX, Verónica de Jesús

Recibido Noviembre 30, 2015; Aceptado Marzo 24, 2016

Resumen

En la Universidad Tecnológica del Suroeste de Guanajuato, se establecieron seis variedades de jitomate (*Lycopersicon esculentum* Mill) tipo bola (Brenda HF1, Retyna HF1, Baptysta HF1, Calvi HF1, Cantyca CV459, Brentyla HF1) bajo condiciones de invernadero y en suelo, para evaluar el efecto del raleo a tres, cuatro y cinco frutos por racimo. Se cosecharon cuatro racimos para evaluar el peso del fruto, diámetro promedio de fruto y rendimiento por planta. Se utilizó el programa estadístico SAS versión 9.2 para un análisis de correlación y una prueba de comparación de medias Tukey ($P \leq 0.05$), para un diseño completo al azar. Hubo correlación positiva entre todas las variables. La variedad Cantyca CV 459 con raleo a 5 frutos presentó el mayor rendimiento (4.1447 Kg) y la variedad Baptysta HF1 con raleo a 3 frutos, tuvo el menor rendimiento (1.33 Kg). La variedad Retyna HF1 con raleo a 4 frutos obtuvo el mayor peso de fruto (281.37 g). La variedad Bentryla HF1 con un raleo a 3 frutos el menor peso (135.3 g). Las variedades Retyna HF1 con raleo a 4 frutos y Baptysta HF1 con raleo a 5 frutos, presentaron el mayor diámetro de fruto.

Calidad de fruto, Raleo de fruto, *Lycopersicon esculentum*

Abstract

Technological University of the Southwest of Guanajuato, six varieties of tomato (*Lycopersicon esculentum* Mill) ball type (Brenda HF1, Retyna HF1, Baptysta HF1, Calvi HF1, Cantyca CV459, Brentyla HF1) were established under greenhouse conditions and ground to assess the effect of thinning to three, four and five fruits per cluster. Four bunches were harvested to evaluate the fruit weight, fruit diameter and average yield per plant. SAS statistical software version 9.2 for correlation analysis and comparison of means test Tukey ($P \leq 0.05$) for a complete randomized design was used. There was a positive correlation between all variables. The Cantyca CV459 variety with fruit thinning 5 had the highest yield (4.1447 Kg) and the variety Baptysta HF1 with 3 fruit thinning, had the lowest yield (1.33 Kg). The Retyna HF1 variety with fruit thinning to 4 had the highest fruit weight (281.37 g). The Bentryla HF1 variety with a fruit thinning to 3 lower weight (135.3 g). Retyna varieties HF1 with 4 fruit thinning and thinning Baptysta HF1 with 5 fruits had the highest fruit diameter.

Quality of fruit, Fruit thinning, *Lycopersicon esculentum*

Citación: GAYTÁN-RUELAS, Marina, VARGAS-ESPINOZA, Everardo, RIVERA-ARREDONDO, Marisa y MORALES-FÉLIX, Verónica de Jesús. Efecto del raleo de frutos en el rendimiento de jitomate (*Lycopersicon esculentum* Mill). Bajo invernadero. Revista de Análisis Cuantitativo y Estadístico. 2016. 3-7: 18-25

*Correspondencia al Autor (Correo electrónico: sergio.martinez@uttt.edu.mx)

† Investigador contribuyendo como primer autor.

Introducción

En los últimos años, la producción tomatera ha aumentado en alrededor de 50 por ciento, impulsada por una mayor superficie agrícola, con sistemas de agricultura protegida. En todos los meses del año se tiene disponibilidad de jitomate (SRE, 2015).

México es el principal exportador de jitomate fresco a nivel mundial, con cerca del 20% del volumen y 25% del valor comercializados, que se destinan principalmente a EEUU. El país exporta alrededor de 1.5 millones de toneladas anuales, que representan entre el 50 y 70% del volumen de producción. En México se siembran alrededor de 52,374.91 ha de jitomate, con un rendimiento promedio de 56.42 t.ha⁻¹, por lo que es la segunda hortaliza más importante en cuanto a superficie sembrada, la más trascendente por su volumen en el mercado nacional y la primera por su valor de producción (SIAP-SAGARPA, 2014).

En la producción de esta hortaliza bajo condiciones protegidas, se busca aprovechar y eficientizar el espacio en vertical de un invernadero; ya que, el microclima generado es controlable para dar un rendimiento óptimo de las plantas, además de favorecer el control de plagas y enfermedades. Por estas y otras razones que competen a la eficiencia productiva; así como a objetivos académicos y de investigación, con fines de transferencia de tecnología hacia los pequeños productores; se decidió implementar el presente proyecto para evaluar el efecto de raleo de frutos bajo tres niveles de raleo, en el rendimiento y calidad de fruto en 6 variedades de jitomate tipo bola, planteando como objetivos el evaluar el efecto de raleo de frutos en el rendimiento y calidad de fruto en seis variedades de jitomate bajo condiciones de invernadero y comparar y determinar el mejor tratamiento de raleo después de la cosecha de cuatro racimos.

Revisión de literatura

Raleo de frutos. La expresión del potencial del rendimiento de los cultivos depende tanto de su constitución genética como de factores ambientales (clima, suelo), factores biológicos y la técnica de producción (Sánchez y Escalante, 1988).

La poda, entendida como la remoción de partes de la planta (yemas, brotes desarrollados, raíces o frutos), sirve para mantener una forma y crecimiento adecuado, siempre y cuando se realice sin afectar el desarrollo de la planta (Halfacre, 1979).

La importancia de la poda radica en que en ocasiones un crecimiento rápido de algún órgano puede competir con las hojas por nutrimentos que fácilmente se pueden translocar, lo que provoca senescencia foliar y reducción en su capacidad fotosintética. Asimismo, existe competencia entre los órganos cuyo crecimiento y desarrollo son simultáneos; tal es el caso del crecimiento del ápice con la diferenciación floral, proceso que ocurre a muy temprana edad en muchas plantas. El crecimiento resultante de una poda es bastante rápido porque se altera, temporalmente, la relación raíz/parte aérea. Además, la remoción de follaje y ramas reduce la cantidad de carbohidratos almacenados y, lo que es aún más importante, reduce el área foliar disponible para su producción (Salisbury y Ross, 1994).

Por otra parte, una poda terminal excesiva estimula el crecimiento vegetativo y puede suprimir la floración, ya que al remover los ápices los meristemas laterales dispondrán de mayor abastecimiento de agua, nitrógeno y otros elementos vitales para el crecimiento vegetativo (Halfacre, 1979).

Así mismo, Sánchez del Castillo, *et al.* (2005), mencionan que la disminución del índice de área foliar permite un aumento en el peso medio de los frutos, pero reduce el rendimiento por unidad de superficie y el raleo de frutos induce un aumento en el peso de los frutos en los racimos, pero el rendimiento por unidad de superficie se aminora.

De acuerdo con Velasco *et al.*, (2011), para conseguir calidad de los frutos, evaluada como el desarrollo óptimo de los frutos y su uniformidad, es muy recomendable eliminar uno, dos o más frutos por racimo, dependiendo el número de frutos amarrados (generalmente los que están en el extremo del racimo, puesto que estos generalmente suelen quedar más chicos y son de poco valor comercial).

Además el tamaño que alcanzan los frutos es influido por la cantidad de fotoasimilados disponibles por planta, pues al eliminar flores los fotoasimilados se reparten entre menos frutos y se propicia un mayor tamaño que ayuda a compensar parcialmente la pérdida de rendimiento por unidad de superficie ocasionada por el menor número de frutos por planta o por racimo (Wolf y Rudich, 1988; Shishido, 1989).

La eliminación de los frutos chicos puede realizarse manualmente o con tijeras, inmediatamente después del amarre o a más tardar cuando los frutos tengan el tamaño de una canica. Ocasionalmente, el raleo de frutos chicos en los racimos también se justifica cuando se presenten síntomas de deficiencia de calcio en los frutos. Velasco *et al.*, (2011).

En las plantas de jitomate tipo bola es conveniente hacer un raleo dejando de tres a cinco frutos por racimo (dependiendo de la variedad) para obtener frutos de buen peso y de tamaño uniforme.

Este raleo también evita que el racimo se caiga antes de que los frutos maduren; factor importante a considerar y que a menudo es un aspecto que se descuida, puesto que la sobre carga de fruto en la planta puede traer graves consecuencias (León, 2006).

Lo esencial del raleo, es buscar un aumento en el calibre del fruto y la formación de yemas, pero a su vez se pueden obtener un sin número de otros beneficios, tales como; un mejor color, un mayor contenido de azúcar, un mejor control sanitario, y entre otros (Salisbury y Ross, 1994).

Sin embargo, es importante señalar que se debe de prestar especial cuidado en las prácticas de polinización y eficiencia de la misma, puesto que de esta depende que se originen frutos con suficientes semillas, al ser las responsables del crecimiento y desarrollo de la pulpa y del fruto mismo. Si no existe una polinización o fecundación óptima la producción disminuye, de allí su importancia (Serrano, 1978).

Ambientes con 60 % de humedad relativa y temperaturas de 26 a 28 °C son idóneos para el proceso de la polinización. La falta de circulación de aire también afecta este proceso; puesto que, al no haber movimiento de las flores el polen no se desprende de las anteras Velasco *et al.*, (2011).

Por otra parte y siendo igual de importante en este proceso, el balance nutricional juega un papel fundamental a tener en cuenta desde el trasplante hasta la etapa productiva del cultivo, razón por la cual es necesario el tomar en cuenta los siguientes aspectos: realizar un análisis previo del estado nutricional del suelo donde se va a sembrar y estimar los índices de extracción de nutrimentos por cosecha.

En este sentido, se calcula que una cosecha promedio de 40,000 kg.ha⁻¹, puede extraer cerca de 180 kg de nitrógeno, 60 a 80 kg de P y 220 kg K. (León, 2006). Dentro de la nutrición del cultivo de jitomate se pueden realizar y aplicar soluciones balanceadas de crecimiento, floración y fructificación o llenado de fruto, en base a pH óptimo de 5.5 a 5.8 y una conductividad eléctrica máxima de 3.5 dS m⁻¹.

Materiales y métodos

El trabajo se realizó durante el periodo de mayo – agosto del 2015 en la Universidad Tecnológica del Suroeste del estado de Guanajuato (UTSOE) en el invernadero 1 de la Carrera de Agricultura Sustentable y Protegida ubicada sobre la Carretera Valle-Huanímaro Km. 1.2 Valle de Santiago, Gto., México.

Material vegetal.

En el invernadero se establecieron 6 variedades tipo bola en suelo, las cuales fueron: Brenda HF1, Retyna HF1, Baptysta HF1, Calvi HF1, Cantyca CV459, Brentyla HF1; manejadas cada una a 3 niveles de raleo: 3, 4, y 5 frutos por racimo; obteniendo un total de 18 tratamientos. Cada tratamiento estuvo constituido por 3 repeticiones; para tener un total de 54 unidades experimentales; cada una evaluada a 4 racimos cosechados en madurez comercial.

Variables a evaluar.

- Rendimiento (Kg/planta⁻¹): se obtuvo cosechando racimos en madurez comercial y pesando con la ayuda de una balanza digital, lo acumulado al cabo de cuatro racimos cosechados por planta.

- Peso de fruto (g): se obtuvo del promedio de todos los frutos de los cuatro racimos cosechados en cada uno de los tratamientos; y se pesó con la ayuda de una balanza digital.

- Diámetro de fruto (cm): se obtuvo del promedio de todos los frutos de los cuatro racimos cosechados en cada uno de los tratamientos; y se midió con la ayuda de un vernier digital.

Manejo agronómico.

En el cultivo se realizaron las prácticas como el deschupone (cada semana); el tutoreo y retutoreo (cada que lo requería el cultivo), las polinizaciones (diariamente a las 10 y 13 hrs). También se realizó la poda de hojas y el raleo de frutos; el mantenimiento de cintilla se hizo cada que se requería; así como el bajado de planta de acuerdo al crecimiento del cultivo.

La nutrición del cultivo se manejó bajo una solución en base a Steiner de 18 a 22 meq, desde el trasplante (15 de abril) hasta cuatro meses después del trasplante.

Se realizaron cuatro riegos por día: dos por la mañana sin solución (10 y 12 hrs), y en las tardes (14 y 16 hrs) con la solución nutritiva anterior.

Para el control de plagas como mosca blanca, trips, ácaro del bronceado y paratrioza; así como para el manejo de enfermedades como damping off, tizón temprano y cenicilla; se realizaron aplicaciones semanales y/o quincenales de plaguicidas orgánicos y químicos en base a dosis recomendadas comercialmente. También se realizaron aplicaciones de micronutrientes foliares cada semana y quincenalmente; así como de aminoácidos.

Análisis de datos.

Se utilizó el programa estadístico SAS versión 9.2 para llevar a cabo la correlación entre las variables evaluadas y una prueba de medias Tukey ($P \leq 0.05$), para un diseño completo al azar con dieciocho tratamientos y tres repeticiones.

Resultados y discusión

En la Tabla 1 se aprecia que existe correlación positiva altamente significativa entre todas las variables (rendimiento por planta, peso y diámetro de fruto).

Por consiguiente, la prueba de comparación de medias Tukey ($P \leq 0.05$) arrojó diferencias significativas entre los tratamientos.

Estas correlaciones suponen que mientras mayor sea el peso y diámetro de los frutos, se tendrá un mayor rendimiento por racimos o por planta, al tener frutos más vigorosos y con mayor contenido de pulpa y agua (Salisbury y Ross, 1994).

De igual manera el análisis demuestra que los frutos más grandes tienden a ser más pesados, debido a que durante su desarrollo pudieron formar semillas suficientes que ayudan a la síntesis de giberelinas para el desarrollo de la placenta y tejidos internos (Serrano, 1978).

	Rto	Diam
Peso	0.52413**	0.93173**
Diam	0.52009**	

Rto: Rendimiento por planta; Diam: Diámetro de fruto. **: significativo con $\alpha \leq 0.01$.

Tabla 1 Coeficientes de correlación entre las variables evaluadas en el experimento de raleo de frutos en seis variedades de jitomate bola

En la Tabla 2 se observa que para la variable rendimiento al cabo de cuatro racimos cosechados, la variedad Cantyca CV 459 con el nivel de raleo a 5 frutos, presentó un mayor rendimiento con 4.14 Kg; dicha variedad fue estadísticamente igual a las variedades Brentyla HF1 R4, Calvi HF1 R3, Retyna HF1 R4, Brentyla HF1 R5 y Cantyca CV459 R4. Contrariamente, la variedad Baptysta HF1 con el nivel de raleo a 3 frutos, fue la que presentó el menor rendimiento (1.33 Kg) bajo estas condiciones de producción y manejo. Lo encontrado para los tratamientos rendidores, se puede atribuir a características genéticas propias de los materiales y que responden bien a la producción y comercialización en racimos de más de 3 frutos por cada uno (Sánchez y Escalante, 1988), al lograr la concentración y optimización de fotoasimilados para el desarrollo óptimo de los frutos (Velasco, *et al.*, 2011).

Contrariamente a lo establecido por Wolf y Rudich, (1988), las variedades Baptysta HF1, Calvi HF1 y Brenda HF1 (con raleo a 3, 4 y 3 frutos respectivamente) no reunieron los fotoasimilados ocasionada por el menor número de frutos por planta o por racimo, para compensar parcialmente la pérdida de rendimiento por unidad de superficie.

Tratamiento	Rendimiento. Pta ⁻¹ (Kg)
Cantyca CV459 R5	4.1447 a
Brentyla HF1 R4	3.9147 ab
Calvi HF1 R 3	3.5960 abc
Retyna HF1 R4	3.4000 abcd
Brentyla HF1 R5	3.2707 abcd
Cantyca CV459 R4	3.2700 abcd
Cantyca CV459 R3	3.058 bcde
Retyna HF1 R5	2.9867 bcde
Brenda HF1 R4	2.8260 cdef
Calvi HF1 R5	2.8253 cdef
Retyna HF1 R3	2.7960 cdef
Brentyla HF1 R3	2.6993 cdef
Baptysta HF1 R5	2.4907 def
Brenda HF1 R5	2.4467 def
Baptysta HF1 R4	2.2157 efg
Brenda HF1 R3	1.9397 fg

Calvi HF1 R4	1.8960 fg
Baptysta HF1 R3	1.3333 g
DMS: Diferencia mínima significativa; *Valores con la misma letra en la misma columna son estadísticamente iguales (Tukey, $\alpha < 0.05$). R3, R4, R5: Nivel de raleo a tres, cuatro y cinco frutos, respectivamente.	

Tabla 2 Efecto de raleo de fruto en el rendimiento de seis variedades de jitomate tipo bola

Respecto a la variable peso de fruto, en la Tabla 3 se observa que la variedad Retyna HF1 con raleo a 4 fue la que mayor peso de fruto obtuvo (281.37 g). Inversamente la variedad Bentryla HF1 con un raleo a 3 frutos tuvo un peso promedio de 135.3 g. Lo anterior pudo generarse por el nivel de raleo, puesto que a pesar de que en las plantas de jitomate tipo bola es conveniente hacer un raleo dejando de tres a cinco frutos por racimo, depende de las características de la variedad para obtener frutos de buen peso y de tamaño uniforme (León, 2006).

Tratamiento	Peso (g)
Retyna HF1 R4	281.37 a*
Baptysta HF1 R5	225.86 ab
Calvi HF1 R 3	221.14 abc
Cantycya CV459 R5	210.54 abcde
Brenda HF1 R5	202.91 abcd
Calvi HF1 R5	187.04 bcdefg
Retyna HF1 R5	186.63 bcdef
Brentyla HF1 R4	180.87 bcdef
Brenda HF1 R4	176.88 bcdefg
Retyna HF1 R3	174.75 bcdefg
Cantycya CV459 R4	166.3 bcdefg
Brentyla HF1 R5	163.53 cdefg
Cantycya CV459 R3	162.17 defg
Brenda HF1 R3	161.79 cdefg
Calvi HF1 R4	158.36 defg
Baptysta HF1 R3	144.22 efg
Baptysta HF1 R4	138.48 fg
Brentyla HF1 R3	135.3 g
DMS: Diferencia mínima significativa; *Valores con la misma letra en la misma columna son estadísticamente iguales (Tukey, $\alpha < 0.05$). R3, R4, R5: Nivel de raleo a tres, cuatro y cinco frutos, respectivamente.	

Tabla 3 Efecto de raleo en el peso de fruto de seis variedades de jitomate tipo bola

Anteriormente se observó que en general la variedad Brentyla HF1 en sus niveles de raleo de frutos 4 y 5, resultó con mayor rendimiento, no obstante cuando se realizó el raleo a 3 frutos, el peso de los mismos se ve disminuido.

Lo anterior se puede deber a que la planta puede llegar a un estrés por pérdida de nutrientes, reflejándose en la disminución del peso del fruto.

Además, la remoción de follaje, frutos y ramas reduce la cantidad de carbohidratos almacenados. (Salisbury y Ross, 1994).

En la variable diámetro de fruto, las variedades Retyna HF1 R4 y Baptysta HF1 R5, presentaron el mayor valor, siendo estadísticamente iguales entre ellas. Estas mismas variedades obtuvieron el mayor peso en ambos niveles de raleo (4 en caso de Retyna HF1 y 5 en caso de Baptysta HF1).

No obstante, para la misma variable (diámetro de fruto), las variedades Baptysta HF1 a 4 frutos raleados y Brentyla HF1 a 3 frutos, presentaron el menor tamaño.

Se observó que la variedad Baptysta HF1, en raleos a 3 frutos, disminuye su peso afectando consecuentemente el rendimiento, atribuido posiblemente a las características de la variedades para obtener frutos de buen peso y de tamaño uniforme (León, 2006), así como de su constitución genética y factores ambientales (clima, suelo), factores biológicos y la técnica de producción (Sánchez y Escalante, 1988).

Tratamiento	Diámetro (cm)
Retyna HF1 R4	8.38 a
Baptysta HF1 R5	7.87 ab
Calvi HF1 R 3	7.76 abc
Brenda HF1 R5	7.60 bcde
Cantycya CV459 R5	7.48 bcd
Brentyla HF1 R4	7.33 bcdef
Retyna HF1 R5	7.30 bcdef
Retyna HF1 R3	7.20 bcdef
Brenda HF1 R4	7.09 bcdef
Calvi HF1 R5	7.03 bcdef
Cantycya CV459 R4	7.01 bcdef
Brenda HF1 R3	6.93 cdef
Brentyla HF1 R5	6.89 cdef
Calvi HF1 R4	6.78 def
Cantycya CV459 R3	6.71 cdef
Baptysta HF1 R3	6.58 ef
Baptysta HF1 R4	6.49 f
Brentyla HF1 R3	6.37 f
DMS: Diferencia mínima significativa; *Valores con la misma letra en la misma columna son estadísticamente iguales (Tukey, $\alpha < 0.05$). R3, R4, R5: Nivel de raleo a tres, cuatro y cinco frutos, respectivamente.	

Tabla 4 Efecto de raleo en el diámetro de fruto de seis variedades de jitomate tipo bola

Conclusiones

Existe correlación positiva altamente significativa entre todas las variables (rendimiento por planta, peso y diámetro de fruto). La variedad Cantycya CV 459 con el nivel de raleo a 5 frutos presentó el mayor rendimiento (1036.17 Kg), así mismo la variedad Bentryla HF1 con el nivel de raleo 4.

La variedad Baptysta HF1 con el nivel de raleo a 3 frutos, tuvo el menor rendimiento (272.99 Kg).

La variedad Retyna HF1 con raleo a 4 frutos obtuvo el mayor peso de fruto (281.37 g). Inversamente la variedad Bentryla HF1 con un raleo a 3 frutos obtuvo el menor peso (135.3 g). Las variedades Retyna HF1 R4 y Baptysta HF1 R5, presentaron el mayor diámetro de fruto.

Estas mismas variedades obtuvieron el mayor peso en ambos niveles de raleo (4 en caso de Retyna HF1 y 5 en caso de Baptysta HF1).

Referencias

HALFACRE G., R. 1979. Horticultura. AGT Editor, S.A. México, D.F. 727 p.

Ho, L. C. (1984). Partitioning of assimilates in fruit tomato plants. *Plant Growth Reg.* 2:277-285.

Hurd, R. G, A. P.; Gay, A. C. Mountfield (1979). The effect of partial flower removal on the relation between root, shoot and fruit growth in the indeterminate tomato. *Ann. Appl. Biol.* 93:77-89.

León, G. H.M. (2006). Guía para el cultivo de tomate en invernadero. Segunda edición. Editorial SEP-INDAUTOR. 263 p. Chihuahua, Chih.

Salisbury B., F.; C. W. Ross. 1994. Fisiología Vegetal. Editorial Iberoamericana. México. 759 p.

Sánchez Del C., F.; E. Escalante, R. 1988. Hidroponía. Ter- 1988. Hidroponía. Tercera edición. Universidad Autónoma Chapingo. Chapingo, México. Pp. 17 y 18.

Sánchez, Del C.; Ucan, Ch. I.; Contreras, M. E.; Corona S. T. (2005). Efecto de la densidad de población y raleo de frutos sobre el rendimiento y tamaño del fruto en tomate. *Revista Fitotecnia Mexicana*. Vol. 28 (1): 33 – 38.

Secretaría de Relaciones Exteriores. (2015). *México: Primer Exportador Mundial de Tomate*. Recuperado de: <http://consulmex.sre.gob.mx/omaha/images/JITOMATE/jitomate.pdf>

Serrano, C. O. Z. (1978). Tomate, Pimiento y Berenjena en Invernadero. Publicaciones de Extensión Agraria. Madrid, España. Pp. 7-159.

Shishido, Y, N; Seyama, S; Imada, Y Hori (1989) Carbon budget in tomato plants as affected by night temperature evaluated by steady state feeding with $^{14}\text{CO}_2$. Ann. Bot. 63:357-367.

SIAP-SAGARPA. (2014).
http://infosiap.siap.gob.mx/aagricola_siap/icultivo/index.jsp

Velasco, H. E.; Nieto, A. R.; Navarro, L. E. (2011). Cultivo del tomate en hidroponía e invernadero. Ed. BBA. México, D.F. 125 p.

Wolf, S. J. Rudich. (1988). The growth rates of fruits on different parts of the tomato plants and the effect of water stress on dry weight accumulation. Sci. Hort. 34:1-11.

Estudio de métodos difusos para la agrupación en un conjunto de datos

FUENTES, Juan Jaime*†

Universidad Politécnica del Valle del Évora, Angostura, Sinaloa, Mexico

Recibido Septiembre 28, 2015; Aceptado Enero 5, 2016

Resumen

En este documento se analiza los resultados de una investigación que realiza una comparación del agrupamiento de las características físicas de semillas de frijol, usando dos métodos de lógica difusa; el primer caso utiliza la lógica difusa con la Fuzzy C-Means algoritmo (FCM), que trata de encontrar similitudes entre las diferentes variables.

Lógica Difusa, Fuzzy C-Means (FM), Sistema de Inferencia Difusa, Agrupación, Frijol

Abstract

This research explain the results in investigation of comparison of a glustering of physical characteristics bean seeds, using two methods of fuzzy logic; the first case uses fuzzy logic with Fuzzy C-Means algorithm (FCM), trying to find similarities between different variables.

Fuzzy Logic, Fuzzy C-Means (FM), Fuzzy Inference System (FIS), Clusters, Beans

Citación: FUENTES, Juan Jaime. Estudio de métodos difusos para la agrupación en un conjunto de datos. Revista de Análisis Cuantitativo y Estadístico. 2016. 3-7: 26-29

*Correspondencia al Autor (Correo Electrónico: juanjaime.fuentes@upve.edu.mx)

† Investigador contribuyendo como primer autor.

Introducción

La agrupación de características es utilizada cuando no existe el conocimiento suficiente sobre los valores de las características de un objeto de estudio. Esto puede ser extendido a la gran mayoría de las ramas de investigación: Biotecnología, Medicina, Ciencias Computacionales, Ciencias Administrativas, etc. El principal objetivo de interactuar con estos datos es clasificarlos en pequeños grupos que describan sus características principales, basándose en la similitud o diferencia entre ellos. Es imposible analizar directamente dicha cantidad de datos, por lo que comúnmente se recurre a la utilización de técnicas de agrupación que ayudan a particionar un conjunto de datos en pequeños grupos que permiten un análisis eficiente de la información. El análisis de grandes volúmenes de datos no sólo puede brindar información adicional, sino también conocimiento nuevo.

Marco teórico

Esencialmente, en un proceso que conduzca a la solución de un problema, este busca analizar los datos disponibles. El análisis de los datos forma el núcleo de la minería de datos, pero el proceso completo abarca también temas tales como la definición del problema y el desarrollo del problema para resolverlo.

El descubrimiento de esta información oculta es posible gracias a la Minería de Datos (del inglés Data Mining), que entre otras sofisticadas técnicas aplica la inteligencia artificial para encontrar patrones y relaciones dentro de los datos permitiendo la creación de modelos, es decir, representaciones abstractas de la realidad, pero es el descubrimiento del conocimiento, KDD, por sus siglas en inglés. (Morrison, 2014) el que se encarga de la preparación de los datos y la interpretación de los resultados obtenidos, los cuales dan un significado a estos patrones encontrados.

Las técnicas de minería de datos son el resultado de un largo proceso de investigación y desarrollo de productos. Esta evolución comenzó cuando los datos fueron almacenados por primera vez en computadoras, y continuó con mejoras en el acceso a los datos, y más recientemente con tecnologías generadas para permitir a los usuarios navegar a través de los datos en tiempo real. La minería de datos toma este proceso de evolución más allá del acceso y navegación retrospectiva de los datos, hacia la entrega de información prospectiva y proactiva. La minería de datos está lista para su aplicación en la comunidad científica porque está basado en tres tecnologías que ya están suficientemente maduras en computación (Zaiane, 2007):

- Recolección masiva de datos.
- Potentes computadoras con multiprocesadores.
- Algoritmos de minería de datos

Para las empresas, el aprovechamiento de la minería de datos de estas características representa retos tanto en infraestructura de almacenamiento y procesamiento como en la captación de personal capacitado que pueda adaptar e innovar para las aplicaciones específicas. Solamente en Estados Unidos de América, se estima que en el año 2018 habrá una necesidad de 140 000 a 190 000 expertos con estos conocimientos (Stackpole, 2012).

Existen algoritmos que apoyan a la creación de agrupaciones y son los de índoles difusas los que más éxito han tenido como lo son: El algoritmo Fuzzy C-Means (FCM) Dumm (1973) pertenece a una clase de algoritmos basados en funciones objetivo, en cambio el algoritmo K-means propuesto por Chih et al. (2011) es un método de agrupamiento, que tiene como objetivo la partición de n datos en k grupos en el que cada grupo pertenece al grupo más cercano a la media.

Desarrollo

Se realizó un estudio para evaluar el comportamiento del algoritmo Fuzzy C-Means(FCM), Fuzzy K-Means(FKM) en la clasificación de semillas de frijol azufrado higuera. La ejecución de los métodos de agrupación difusa se realizó tomando conjunto de datos reales obtenidos de manera propia a través de las características propuestas por Celis (2008) la cual se presenta en la tabla 1.

Variable
Índice cotiledón/testa índice
Cotiledón/eje embrionario
Intensidad (del color)
Luminosidad (del color)
Peso volumétrico Peso de 100 Semillas
Porcentaje de cotiledón
Porcentaje de testa
Porcentaje de eje embrionario
Tono (del color)

Tabla 1 Características generales de una semilla de frijol azufrado higuera

Los cuales son utilizados para el entrenamiento de los algoritmos, estos datos constan de 100 semillas de frijol azufrado higuera de la cuales se clasifican en dos tipos óptimo y no óptimo teniendo como resultado muestra 52 semillas con un resultado óptimo y 48 no óptimo, para obtener la partición de dos grupos.

Se compararon los resultados con respecto a los mismos experimentos. La evaluación de los resultados obtenidos en el estudio de los diferentes métodos difusos; se evalúa de la siguiente manera. Suponiendo que el número final de clasificación es k , la medida de exactitud r está dada por la ecuación siguiente:

$$r = \frac{\sum_{i=1}^k a_i}{n} \quad (1)$$

Donde n es el número de instancias del conjunto de datos, a_i es el número de instancias que aparecen clasificadas correctamente en i y en su correspondiente clase, la cual es aquella que tenga el número máximo. En otras palabras, a_i es el número de instancias con las etiquetas de la clase que dominan en la clasificación i . Por lo tanto, la prueba del algoritmo se realizó con diferentes números de atributos, en los cuales cada caso se evalúa con la ecuación siguiente; para el desarrollo de estos algoritmos fue utilizada la herramienta computacional de Matlab®.

$$E = 1 - r \quad (2)$$

Implementación y resultados

Se realizó un número de entrenamientos que permitieron obtener diferentes resultados, siendo el conjunto de datos utilizados para el algoritmo FCM el mejor resultado fue obtenido con el entrenamiento de las variables Índice Cotiledón, Luminosidad, Peso Volumétrico, Tono. Con el que se obtiene un Error del 0.02, que en números porcentuales estamos diciendo que tenemos un 99,98% de clasificación correcta de los 100 datos entrenados los cuales se muestran en la tabla 2.

Agrupación	Optimo	No optimo	Instancias
1	1	47	48
2	51	1	52
Total	52	48	100
Error =	0.02		

Tabla 2 Entrenamiento realizado con los Datos Índice Cotiledón, Luminosidad, Peso Volumétrico, Tono Algoritmo Fuzzy C-Means.

Para el entrenamiento del Algoritmos FKM del conjunto de los 100 datos utilizados, al igual que con el algoritmo anterior el mejor resultado obtenido fue utilizando las variables Índice Cotiledón, Luminosidad, Peso Volumétrico, Tono.

El error que presenta este algoritmo es del 0.84, que en números porcentuales estamos diciendo que tenemos un 99,16% de clasificación correcta como se muestran en la tabla 3.

Agrupación	Optimo	No optimo	Instancias
1	8	48	48
2	44	8	52
Total	52	48	100
Error =	0.84		

Tabla 3 Entrenamiento realizado con los Datos Índice Cotiledón, Luminosidad, Peso Volumétrico, Tono Algoritmo Fuzzy K-Means

Conclusiones

Los resultados del análisis para el conjunto de datos de las semillas de frijol azufrado higuera, De manera individual lo obtuvo el algoritmo (FCM), de todos los entrenamientos el que obtiene el mejor error promedio para este conjunto de datos es el algoritmo (FKM) como se muestra en el Tabla 4. Los algoritmos propuestos presentan buen desempeño para los conjuntos de datos evaluados en este trabajo. En general, los algoritmos en la mayoría de las pruebas evaluadas, presentan un comportamiento estable en la medición del error. Por lo tanto, esto implica que los resultados en los procesos de agrupamiento mantienen estable la variabilidad de los datos dentro de los conjuntos de datos.

Método	Error	%	Error Promedio
FCM	0.02	99.98%	0.18434
FKM	0.84	99.16%	0.11326

Tabla 4 Resultados obtenidos del procesamiento de datos con los Algoritmos Fuzzy C-Means, Fuzzy K-Means

Recomendaciones

Se puede mejorar el mecanismo para el manejo de datos anómalos en el conjunto de datos, de forma que se encuentren nuevas soluciones para asignar o eliminar estos datos de los agrupamientos.

Otra recomendación que se puede hacer es la creación de algún algoritmo de clasificación difusa que pueda trabajar con datos mixtos y no solo con numéricos de esta manera incrementar la complejidad de la búsqueda de nueva información.

Referencias

Celis Velázquez, R. Características morfológicas y fisiológicas de la semilla de frijol (*Phaseolus vulgaris* L.) domesticado y silvestre y su relación con el desarrollo y establecimiento de la plántula, (Tesis doctoral inédita), Colegio de Postgraduados, 2008

Chih T.C; Jim Z. C; Mu-der J., A Fuzzy K-means Clustering Algorithm Using Cluster Center Displacement, journal of information science and engineering No. 27, 2011, 995-1009.

Dunn, J. C. A Fuzzy Relative of the ISODATA Process and Its Use in Detecting Compact WellSeparated Clusters, Cybernetics and Systems, Vol 3 No. 3, 1973, 32-57

Morrison, D. (2014). Phylogenetic networks: a new form of multivariate data summary for data mining and exploratory data analysis. Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery, 4(4), pp.296-312.

Stackpole, B. Your Big Data To-Do List. Computer World, Feb. 13 2012.

Yang, M.-S. A Survey of Fuzzy Clustering. Mathematical and Computer Modelling, Vol.18, No 11, 1993, 1-16.

Zaiane, O. Principles of Knowledge Discovery in Databases. University of Alberta. Department of Computing Science. 2007

Evaluación del desempeño docente en la LMS Chamilo, de las ingenierías ejecutivas de la Universidad Tecnológica Fidel Velázquez: acciones de profesionalización

HERNÁNDEZ-CRUZ, María Guadalupe*†, ROMERO-ROJAS, Ruth Marcela, ORTIZ-ARANGO, Víctor Tonatiuh y VARGAS-GUTIÉRREZ, Luis Daniel

Universidad Tecnológica Fidel Velázquez

Recibido Diciembre 18, 2015; Aceptado Marzo 15, 2016

Resumen

Este trabajo, tuvo por objetivo evaluar el desempeño docente en la LMS Chamilo de las ingenierías ejecutivas de la Universidad Tecnológica Fidel Velázquez, para ello se establecieron parámetros de ejecución cuantitativa medidas por el número de horas de conexión, número de herramientas usadas en la LMS e interacciones entre alumnos-docente y alumno-alumno como parte de las actividades diseñadas para la enseñanza-aprendizaje. Cualitativamente se diseñó una lista de cotejo que evalúo los contenidos de las actividades disponibles en la LMS. El estudio se realizó en una muestra de 15 docentes de la ingeniería ejecutiva y entre los hallazgos más significativos están el número reducido de herramientas usadas y la poca interacción virtual entre alumnos-docentes y alumno-alumno, cualitativamente se identifica que el diseño instruccional de las actividades suele ser deficiente. Este trabajo, fue la base para el diseño de un curso de capacitación en competencias digitales.

Evaluación docente, LMS Chamilo, Profesionalización

Abstract

This work has the objective, take value of the effort of the L.M.S. Chamilo of the executive Engineers of the Technologic University Fidel Velazquez, Which had established some parameters of execution where quantitative measurements for the number of hours in connexion and allsow number of tools used in the L.M.S. and interactions between students and docents like part of the activities designed for the knowledge of teaching and learning between these. This list was designed o evaluate the contents of the remaning activities fot the L.M.S.

The study accured as a presentation for 15 docents of the executive Engeneering and in between the most significant discoveries there was a reduced number of tools used in the interaction between students and docents, which is identified in the instructional designee of the activities that usually is deficient.

This work was the beginning of a designee of a course of training of digital competitions.

Evaluation of teachers, LMS Chamilo, Profesionalitation

Citación: HERNÁNDEZ-CRUZ, María Guadalupe, ROMERO-ROJAS, Ruth Marcela, ORTIZ-ARANGO, Víctor Tonatiuh y VARGAS-GUTIÉRREZ, Luis Daniel. Evaluación del desempeño docente en la LMS Chamilo, de las ingenierías ejecutivas de la Universidad Tecnológica Fidel Velázquez: acciones de profesionalización. Revista de Análisis Cuantitativo y Estadístico. 2016. 3-7: 30-44

*Correspondencia al Autor (Correo Electrónico: lupis_hc@hotmail.com)

† Investigador contribuyendo como primer autor.

Introducción

Las ingenierías en modalidad ejecutiva (semipresencial) ha sido impartidas en la UTFV desde el año 2010, dicho modelo cada ciclo cuatrimestral aumenta el número de alumnos que se matriculan en ella y que prefieren este modelo pues les permite incorporarse a la vida laboral o continuar con ésta y concluir el ciclo formativo de ingeniería. De este modo, el perfil del alumno se supone distinto, al del que se encuentra matriculado en el modelo escolarizado, pues la experiencia profesional derivada del trabajo que ya desempeña en el ramo que se encuentra formándose, enriquece el proceso de enseñanza – aprendizaje. Sin embargo al ser un modelo relativamente nuevo en la UTFV la eficiencia y calidad de los procesos y sus actores implicados, ha sido poco abordada y actualmente resulta imperioso medir el éxito que se ha tenido e identificar las áreas de oportunidad que se encuentren en éste, a fin de proponer, diseñar e implementar acciones de mejora.

En México, se gesta un momento histórico donde en el rubro de educación se ha observado la necesidad de evaluar el desempeño docente, creemos fundamentalmente que es necesario, pero no factor único que dentro del proceso educativo deba evaluarse, por lo tanto desde una perspectiva sistémica ha de medirse, la infraestructura, las políticas y gestiones de la institución, el desempeño docente (la actuación y los contenidos) así como los alumnos (rendimiento y actitudes) y solo de este modo podrá tenerse una perspectiva que nos dirija a la mejora y con ello a elevar la calidad educativa y las competencias profesionales con las que egresa un estudiante de educación superior en la UTFV, por lo tanto en el diseño de la evaluación debieran participar actores provenientes de todos los niveles implicados y es parte de la riqueza que se supone tiene el equipo que desarrolla esta línea de investigación.

Para evaluar el desempeño docente en las ingenierías ejecutivas de la UTFV, fue necesario fragmentar el trabajo y diferenciarlo en uno de los elementos distintivos del modelo, el docente ejecuta la mayor parte de su trabajo en una plataforma LMS y lo combina con la práctica docente presencial, pero por la complejidad de ello, se decidió centrar esta evaluación en la ejecución docente solo en la plataforma LMS Chamilo para posteriormente diseñar acciones de profesionalización que resuelvan las problemáticas que se identifican a través de ésta.

Marco teórico

Las TICs en la educación

El incorporar la tecnología y comunicación a la educación proporciona fortaleza que permite transformar aspectos en el quehacer educativo. El uso de recursos didácticos apoyados en la tecnología, a la práctica docente, permite incrementar los servicios habituales que una institución educativa puede brindar. Al ser aplicados en el aula dan un giro a la educación tradicional.

Históricamente hablando, la incorporación de recursos didácticos apoyados en TICs, se han generado como producto del propio avance tecnológico y se han adoptado con bastante facilidad en el ámbito educativo, en el que han encontrado aquí un nicho digno de ser explotado, con todas sus bondades, tal es el caso de las computadoras, calculadoras, sensores, laboratorios virtuales, software de aplicación especializado por áreas del conocimiento, traductores, simuladores, actualmente dispositivos móviles, haciéndose dichas herramientas indispensables para docentes y alumnos se han apropiado de estos recursos mejoran su actividad en los procesos de enseñanza y aprendizaje.

No hay que perder de vista el papel que sigue desempeñando el docente, quien se convierte en guía, tutor o mentor fomentando un ambiente de exploración en clase y en el que el alumno es un actor activo, también responsable de su propio aprendizaje. Todavía es un campo poco explorado con recursos ilimitados.

En el ámbito de la educación superior, las T.I. (Tecnologías de Información) permiten preparar a los estudiantes para un mercado laboral que utiliza de manera exhaustiva la tecnología como instrumento de innumerables procesos.

Este escenario es un reto para las universidades, que tienen que asegurarse de acercar dicha tecnología lo más posible tanto a docentes como alumnos, esto se reflejaría en un aprendizaje de mayor calidad al dotar a los estudiantes de habilidades pertinentes al entorno actual y este es un reto permanente y cambiante por lo que el uso de la tecnología se vuelve imprescindible.

Muchos de los avances en distintas ramas, se originaron a raíz de la Segunda Guerra Mundial tales como técnicas de planeación en áreas administrativas, de control, administración de proyectos, tecnología.

La Tecnología Educativa surge en los años 40's ante la necesidad de formar en corto tiempo a un gran número de ciudadanos, en soldados y oficiales capacitados en acciones bélicas.

Educadores y psicólogos pusieron en práctica programas de formación con el apoyo de recursos audiovisuales, midiendo los resultados del aprendizaje a través de pruebas estandarizadas. (Área Moreira, 2009)

Durante los 50's y 60's la Tecnología Educativa toma un nuevo rumbo, más enfocado a la difusión y el impacto social de las masas a través del radio, cine, prensa y T.V.

Se estudia el aprendizaje del ser humano bajo parámetros de la psicología conductista. En los 70's se alcanza una época de esplendor, se consolidan asociaciones y academias profesionales, también se publican gran cantidad de revistas especializadas en el tema.

En este periodo cobra relevancia el análisis y descripción de las tareas como eje de diseño tecnológico de la enseñanza, (Groppe 1980, en: Área Moreira. 2009) programada como la máxima expresión del diseño tecnológico aplicado a la acción formativa.

En los 80's y 90's se tiene una crisis de la perspectiva tecnócrata sobre la enseñanza. Se cuestiona el uso de la Tecnología Educativa y se llega a considerar al docente y también a las instituciones de educación solo consumidores de contenidos. Al comienzo del siglo XXI, el propio avance de la tecnología permite reformular varias tesis.

En la actualidad existen muchas definiciones sobre la Tecnología Educativa, para Área Moreira (2009), es una disciplina que estudia los procesos de enseñanza y de transmisión de la cultura mediados tecnológicamente en distintos contextos educativos.

El ámbito de estudio de la Tecnología Educativa son las relaciones entre las tecnologías y la educación.

Modalidades de la educación a distancia y semipresencial

Jaime Sarramona (1991) en García Aretio (2001) define la Educación a Distancia (EaD) como la metodología de enseñanza donde las tareas docentes acontecen en un contexto distinto de las discentes, de modo que éstas resultan, respecto de las primeras, diferidas en el tiempo, en el espacio o en ambas dimensiones a la vez.

Dentro de esa misma evolución de la Educación a Distancia, surgió una modalidad conocida como Educación semipresencial, este es un modelo en el que realiza una prudente combinación de las técnicas de educación presencial (tutorías, seminarios, coloquios, convivencias, etc.) con las técnicas de educación a distancia (unidades didácticas, cuadernillos de pruebas de evaluación, grabaciones radiofónicas o audiovisuales, etc.) (Oliver Domingo, 1995)

LMS

Es el nombre técnico de una plataforma, en español suele llamársele AVA (Ambiente Virtual de Aprendizaje), sistema de gestión de cursos o LMS por sus siglas en inglés (González y Hernández, 2011).

Con poco más de 20 años de historia, las llamadas plataformas LMS (Learning Management System), han evolucionado hasta convertirse en sistemas a nivel empresarial comprados de manera institucional. Indica también un cambio en la forma en el uso de LMS al incorporarse al quehacer diario de enseñanza y aprendizaje.

Las distintas modalidades de Educación a Distancias (EaD) tienen una base de infraestructura en las tecnologías de la información que a su vez soportan un sistema que es la parte central de la gestión de las actividades de aprendizaje en entornos virtuales, es decir, es un software que permite que todos los profesores y estudiantes cuenten con un espacio de recursos e interacciones en línea al servicio de la organización (González y Hernández, 2011).

LMS: Chamilo

Chamilo LMS o Chamilo Learning Management System es un software que permite crear un campus virtual para la impartición de formación online o semi-presencial.

El proyecto Chamilo inicia en Enero de 2010, aparece como fork o derivación del proyecto Dokeos, un proyecto surgido en el 2004 a su vez surgido a raíz de Claroline, como parte de llevar estas plataformas a entornos empresariales y sacarlo del marco puramente universitario.

Chamilo es una plataforma de e-learning y colaboración, presentada como un software de código abierto, que puede ser adaptado a proyectos educativos o profesionales.

Se distribuye bajo la licencia GNU/GPLv3, que están obligados a aceptar cuando se usa la plataforma.

El término de Chamilo surge de una derivación musicalmente atractiva del término inglés chameleon, o camaleón, representativa de la alta adaptabilidad de Chamilo LMS a sus necesidades de aprendizaje como de trabajo en equipo.

Los docentes que utilizan Chamilo se benefician de una serie de herramientas que le permiten crear un ambiente de aprendizaje eficiente en un tiempo razonable, que es una razón fundamental por la que la UTFV lo usa.

La UTFV a través de las direcciones académicas de los programas de ingeniería, así como el departamento gestor de ellas, ha considerado que sus características la hacen idónea por encima de otros LMS, por lo que es la base de la modalidad mixta también denominada Ejecutiva, que tiene ya algunos años de ofertar.

Durante la creación de un curso, el docente puede:

- Importar o crear documentos (audio, vídeo, imágenes) y publicarlos.
- Construir ejercicios, exámenes y auto-evaluaciones.

- Elaborar o importar lecciones (SCORM y AICC).
- Permitir la entrega de trabajos virtuales.
- Describir los componentes del curso a través de secciones descripción.
- Comunicarse a través del foro o chat, publicar anuncios, agregar enlaces.
- Crear grupos de trabajo.
- Participar en el aula virtual (a través de la extensión videoconferencia).
- Gestión de resultados a través de la herramienta de evaluación.
- Crear encuestas, añadir una wiki para crear documentos con aportes colaborativos, usar un glosario, usar una agenda, gestionar un proyecto.
- Hacer el seguimiento del curso por alumno y guardar copias de los mismos.
- Registrar asistencias.
- Elaborar un diario de clase (progreso del curso).

Chamilo es una plataforma flexible en la que todas sus herramientas se pueden personalizar según las necesidades de cada curso. Este sistema ofrece un intuitivo control de herramientas y administración de espacios, ya que no requiere ninguna habilidad técnica avanzada.

Características del modelo educativo semipresencial de las ingenierías ejecutivas de la UTFV

El proceso de enseñanza-aprendizaje se caracteriza de 7° a 11° cuatrimestre por una formación académica 50% presencial los viernes de 18:00 a 22:00 horas y los sábados de 7:00 a 15:00 horas en las instalaciones universidad y 50% no presencial, la formación no presencial se realiza mediante el uso de la plataforma de educación a distancia.

El Modelo Ejecutivo de Enseñanza, está basado en el paradigma constructivista de la educación, dando especial énfasis a la construcción de aplicaciones del conocimiento científico o técnico; con el apoyo de recursos virtuales, es consistente y complementario al Modelo Basado en Competencias, ya que permite a través del uso de la Internet realizar actividades lúdicas de aprendizaje definidas dentro de los programas de estudio.

Es decir, la práctica en lugar de ser llevada a cabo a través de laboratorios físicos, se realiza a través de simuladores cuyo acceso es a través de la RED. En términos resumidos, el Modelo Ejecutivo de Enseñanza combina horas de enseñanza presencial, con práctica asistida vía Internet (disponible en: <http://www.utfv.edu.mx> 10/06/16)

El docente en la educación semipresencial

La educación semipresencial también denominada "Formación Combinada" o "Aprendizaje Mezclado" es una modalidad de estudios que incluye tanto formación virtual como presencial, como indica Acosta (2009) "el termino blended learning apareció en 2002, que se traduce literalmente como aprendizaje mixto, es decir, esta modalidad pretende utilizar dos estrategias, la presencial y la virtual, por lo tanto, selecciona lo mejor de las dos".

La formación virtual o e-learning requiere cambios profundos tanto en el rol del profesor como en el del alumno. El papel del profesor pasa de transmisor de conocimientos a tutor o guía del proceso de aprendizaje de los estudiantes; asimismo, las TICs en las diferentes modalidades de enseñanza/aprendizaje adquieren una importancia fundamental como diseñador de materiales, medios y estrategias mediadas de aprendizaje, creando en el entorno las situaciones propicias para ello.

La figura del tutor es de suma relevancia en cualquier modalidad formativa (cursos presenciales, semipresencial o en línea), pues es el proceso que permite al profesor organizar, dinamizar, apoyar al alumno en sus aprendizajes. Si bien estos procesos no son específicos de la enseñanza on-line, adquieren en este contexto mediado una significación o importancia mayor, dado el carácter de "individualidad" que imprime el medio electrónico.

Una de las funciones más importantes que tiene un tutor on-line es ser un dinamizador social., por el carácter social del proceso de formación en la enseñanza on-line debe ser tenido en cuenta para crear un clima de trabajo óptimo y un nivel de satisfacción y de pertenencia al grupo en los participantes.

Existen diferentes recursos que pueden ayudar en este proceso, algunos ejemplos pueden ser:

- Añadir la foto de todos los participantes para lograr una mayor cercanía y conocimiento de los mismos.
- Propiciar a través de chat un mayor conocimiento entre los participantes (intereses, motivaciones, etc.).
- Crear un foro de alumnos, en donde tengan libertad para expresar sus ideas. Incluir tareas de trabajo colaborativo que fomenten la interacción entre participantes.
- Seguimiento del nivel de actividad (grupal e individual).

Uno de los grandes problemas con los que se enfrenta la enseñanza a distancia es la deserción de los participantes, lo cual puede deberse a diferentes causas: falta de tiempo, falta de motivación, dificultad para comprender los contenidos, causas personales, etc.

El tutor deberá hacer un esfuerzo importante para realizar el seguimiento tanto del grupo como a nivel personal intentando detectar los posibles problemas que van surgiendo y realizando las acciones necesarias para solucionarlos, poniéndose en contacto con los interesados a través del correo electrónico.

Apoyo técnico. Dado el carácter mediado del proceso de aprendizaje el tutor debe de conocer y atender a los problemas que puedan encontrarse los estudiantes en el proceso de aprendizaje relativos al uso del entorno virtual. Una de las posibilidades es el uso de un foro específico para este tipo de problemas.

Apoyo administrativo. El tutor debe facilitar al estudiante toda la información y apoyo necesario para que pueda realizar las gestiones administrativas relacionadas con su curso o titulación. Para ello, deberá estar al tanto del proceso que se ha seguido y resolver las dudas o problemas que se puedan plantear.

Facilitador y dinamizador del aprendizaje. Es, sin lugar a duda, el rol fundamental del tutor. La transmisión de conocimientos eje fundamental de la enseñanza presencial, ocupa un lugar secundario en el aprendizaje on-line en la que el profesor-tutor se convierte en una figura de apoyo.

Cuando hablamos de competencias, estamos haciendo referencia a conocimientos, habilidades y actitudes que los profesionales han de poseer para resolver problemas en contextos reales. Como señalan Martinet, Raymond y Gauthier (2001), una competencia es siempre una competencia para la acción. Según estos autores:

- La competencia se desarrolla en contextos profesionales reales, no simulados.
- La competencia se sitúa en un continuo que va de lo simple a lo complejo.

- La competencia se fundamenta en un conjunto de recursos: la persona competente hace uso de recursos que moviliza en contextos de acción.
- La competencia tiene que ver con la capacidad de movilizar en contexto de acción profesional.
- La competencia, como saber - actuar es una práctica intencionada.
- La competencia es un saber - actuar eficaz, eficiente e inmediato que se manifiesta de manera recurrente.
- La competencia constituye un proyecto.

Después del análisis es conveniente agrupar las competencias necesarias en e-learning en cuatro dimensiones:

a) Competencias tecnológicas.

Son aquellas habilidades para gestionar y utilizar todos los recursos tecnológicos necesarios para el diseño y desarrollo de la tele formación desde un punto de vista técnico (Internet, herramientas síncronas y asíncronas).

b) Competencias de diseño.

Es aquella habilidad para aplicar los principios didácticos y pedagógicos para el diseño instruccional de lo que es la planificación, desarrollo y evaluación de la acción formativa, con el objeto de crear objetos de aprendizaje que den respuesta a las necesidades del alumnado.

c) Competencias tutoriales.

Se deben poseer destrezas en el manejo de herramientas comunicativas síncronas y asíncronas. Las primeras serán para la comunicación en tiempo real de manera simultánea (chat, videoconferencia), por el contrario en las asíncronas no se comparte la misma estructura temporal (foros, correo electrónico).

d) Competencias de gestión.

Se refieren a poseer conocimientos y habilidades para la coordinación de equipos de trabajo, organización, liderazgo, y en general el óptimo funcionamiento de los recursos humanos en torno a la acción de tele formación.

La planeación de los contenidos en la educación semipresencial

Planificador y organizador. En la formación virtual los tiempos destinados a las diferentes tareas que deben de realizar los estudiantes debe ser planificada previamente al desarrollo del mismo, esto conlleva un análisis pormenorizado de las tareas y tiempos necesarios para su ejecución, así como tener presente los diferentes ritmos que pueden llevar los participantes en el curso lo que conlleva una mayor flexibilidad que en la enseñanza presencial.

La evaluación del desempeño docente

La necesidad de evaluar los resultados del proceso educativo en todos los niveles, necesariamente incluye a los docentes, sin embargo el momento histórico actual, parece dirigirnos a las evaluaciones punitivas para responsabilizar al profesor de las fallas o carencias.

Loredo (2004) plantea que la evaluación docente implica una filosofía, una postura epistemológica, un método y no solo una aplicación de instrumentos para describir o comparar resultados o indicadores.

La complejidad que supone las diversas dimensiones de la labor docente, no solo obligan a pensar las técnicas y métodos con los cuales se evaluará, sino el uso estratégico de estos recursos (Canales, 2004)

Quizá uno de los primeros aspectos que tengan que resolverse, sea ¿Cuál es el objetivo de la evaluación? Puesto que puede caerse en el error de evaluar planes, programas, rendimiento académico y otros aspectos que no son propios de la evaluación del desempeño docente, aun cuando son parte del proceso educativo.

Valdés (2007) distingue cuatro funciones básicas de la evaluación del desempeño docente: la función diagnóstica, para conocer aciertos y desaciertos del profesor con la intención de derivar acciones de formación y superación docente. La función instructiva, para que los profesores se instruyan y aprendan de sí mismo e incorporen nuevas experiencias de aprendizaje profesional. La función educativa, donde una vez que el profesor conoce como otros profesores, sus alumnos o las autoridades perciben su trabajo, puede trazar estrategias para superar las debilidades señaladas y finalmente la función de desarrollo profesional del docente, cuando el profesor se torna capaz de autoevaluar de manera crítica y permanente su desempeño no teme sus desaciertos más bien aprende de ellos y se espera que de su insatisfacción surja la necesidad de avanzar en su desarrollo profesional.

Para Monroy (2009) es necesario y fundamental delimitar los componentes de lo que se desea evaluar, es por ello que se ha delimitado el rendimiento docente en este primer ejercicio solo el uso que el docente da a la LMS Chamilo, sabiendo que de esta forma existen muchos aspectos que no se están evaluando, sin embargo en futuras investigaciones pretende abordarse, pero tal como lo plantea Valdés (2007) es necesario que el profesor tenga la percepción de que la evaluación no ponen en riesgo su supervivencia laboral y profesional para generar una evaluación objetiva, pues de otro modo el riesgo es que el profesor tienda a comportarse como cree que se espera independientemente de sus convicciones educativas o la riqueza profesional que pudiera generarse.

Por lo que al ser esta evaluación planeada y dirigida por un equipo de docentes y alumnos colaborando para la investigación y no una acción institucional, esperamos lograr el objetivo que Monroy (2009) cita, que espera que la evaluación docente sea auténtica, justa y racional, que permita valorar el desempeño docente con objetividad, profundidad e imparcialidad. Agregando que sea de utilidad para la constante mejora de la calidad del proceso educativo de forma integral

Método

Sujetos

Se determinó una muestra no probabilista por el método de cuotas de 15 profesores que hubieran impartido clases en las ingenierías ejecutivas de la UTFV, en los últimos tres cuatrimestres continuos al momento del estudio.

Aparatos y materiales

Computadora.

Software SPSS (Statistical Package for Social Science) para Windows V.22 en español

Lápices, plumas y hojas.

Instrumentos

Carta de consentimiento informado.- en la cual se especifica el tipo de información solicitada y garantizaba la confidencialidad del nombre y datos personales del profesor así como la no asociación de los datos derivados de la evaluación a su identidad dentro de la investigación y fuera de ella que ejercicio de evaluación no fue sustitución de la evaluación docente registrada en el sistema de gestión de calidad (SGC) ni los resultados reconocidos como una medida de evaluación sino como un ejercicio de investigación para diagnosticar y planear acciones de profesionalización.

Hoja de registro.- para evaluar y registrar el número de herramientas que cada sujeto de investigación utiliza en la LMS de las ingenierías ejecutivas de la UTFV.

Lista de cotejo.- para la evaluación cualitativa de actividades que los sujetos de investigación colocaron en la LMS.

Diseño

Se realizó una investigación descriptiva y exploratoria, cuasi experimental.

Procedimiento

Se determinó la muestra una muestra no probabilística por el método de cuotas de 15 profesores que hubieran impartido clases en la modalidad de ingenierías ejecutivas de la UTFV en los cuatrimestres 2014-3, 2015-1, 2015-2 y que tuviera asignación de materias en la modalidad, en el cuatrimestre 2015-3. Para conseguir la muestra se convocó a todos los profesores que imparten clases en las ingenierías ejecutivas, se explicó el protocolo de investigación así como los alcances de la carta de consentimiento informado, al finalizar la información los primeros 15 docentes que firmaron la carta, fueron considerados la muestra definitiva de este estudio.

Para garantizar la protección de la identidad de la muestra, su nombre fue sustituido por un número consecutivo que solo conoció el administrador de la LMS y uno de los miembros del equipo de investigadores, posteriormente cada sujeto fue denominado de acuerdo a su número consecutivo.

El administrador de la LMS, proporciono los datos necesarios para el llenado del formato de registro, así como la extracción de alguna de las actividades (de manera textual como aparecía en la plataforma), una vez que se concluyó el mismo procedimiento para los 15 casos.

El análisis de los datos se realizó de la forma siguiente: se categorizaron los datos obtenidos en el formato de registro para ser ordenados usando la estadística descriptiva usando el software SPSS V.22 para Windows.

En el caso de las actividades para el análisis cualitativo, al azar se eligieron solo 4 de entre las quince actividades, las cuales fueron entregadas a tres profesores que imparten la misma materia pero en la modalidad presencial, bajo la metodología de testigo doble ciego, es decir, los tres profesores jamás conocieron el nombre del autor de la actividad y el autor no conoció que actividad fue extraída de la LMS, ni quienes fueron los profesores expertos que la evaluaron, a fin de garantizar la objetividad en el proceso. Cada profesor evaluador, recibo una lista de cotejo con diez puntos a calificar en escala Likert entre muy adecuada y completamente inadecuada y una sección abierta para comentarios, el análisis de los datos de la lista de cotejo se realizó con el reporte de frecuencias y el análisis de los comentarios por categorías.

Resultados

De la muestra de 15 docentes, obtenida por el método de cuotas, estuvo compuesta por seis mujeres (56%) y nueve varones (44%) con una edad promedio de 41.8 años. Los años promedio de experiencia docente en educación superior fue de 12.6 años, mientras que la media fue de 2.3 años de experiencia en impartición de clases en la modalidad semipresencial de las ingenierías ejecutivas de la UTFV. La muestra de profesores reciben en promedio 14.6 horas a la semana de horas/clase en la modalidad ejecutiva (semipresencial), lo que representa entre el 23.5% al 34.8% de la carga total de horas asignadas al cuatrimestre. Ninguno de los docentes que conformaron la muestra imparte exclusivamente clases en la modalidad, sino que la combina con la impartición de clases presenciales en la UTFV.

El primer elemento evaluado, fue la correspondencia entre el número de materias y horas asignadas a cada caso de la muestra y el número de horas que la LMS registró que el profesor estuvo conectado, durante el cuatrimestre 2015-3 (en cuatro semanas que fueron elegidas al azar), observando que el 100% de la muestra (ver, Gráfico 1) permaneció conectado el total de número de horas esperadas, según su asignación de materias.

Correspondencia de horas de conexión

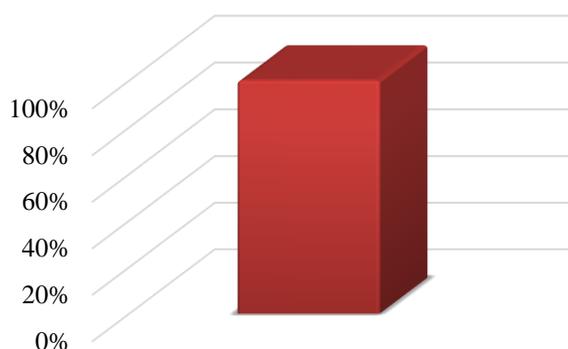


Gráfico 1 Correspondencia entre el número de horas asignadas por profesor y número de horas de conexión en plataforma en cuatrimestre 2015-3

Bajo la misma mecánica de elegir cuatro semanas al azar, de las comprendidas en el periodo 2015-3, se contabilizaron el número de actividades, por semana/materia (ver, Gráfico 2), cada docente indicó para realizar a través de la LMS, encontrando que el 24% de muestra subió dos actividades por materia/semana, el 62% colocó en la LMS tres actividades por semana/materia, mientras que el 14% registró en promedio 5 actividades por semana/materia.

Actividades en plataforma

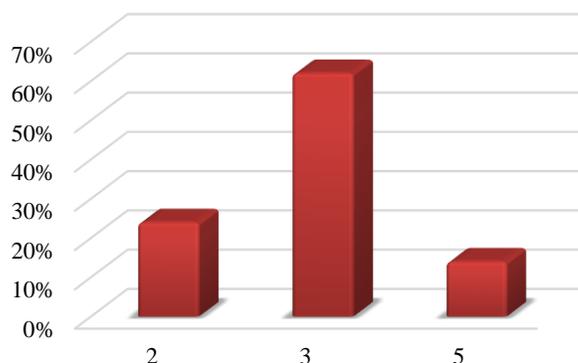


Gráfico 2 Promedio de número de actividades en plataforma por profesor / materia por semana, cuatrimestre 2015-3

El número de interacciones entre el docente y alumno a través de la LMS, evaluada en cuatro semanas al azar en el periodo 2015-3 por grupo/materia (ver, Gráfico 3); el 9% de los docentes mantuvieron en promedio ocho interacciones, el 39% de la muestra tuvo 13 interacciones, mientras que el 52% tuvo en promedio 15 interacciones por semana/grupo, esto en resolución de dudas, tutorías sobre actividades o retroalimentación sobre ejercicios.

Interacción docentes a alumnos

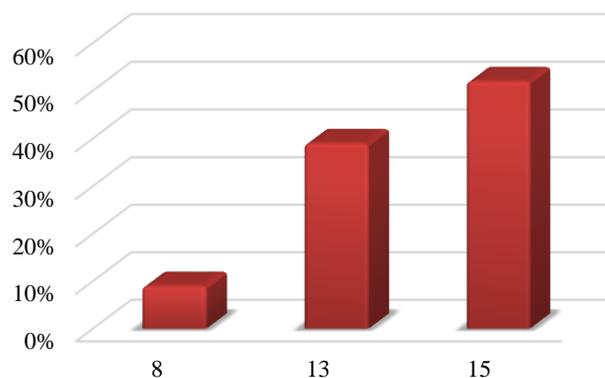


Gráfico 3 Promedio de número de interacciones (respuestas, comentarios/retroalimentación) de profesor a alumno por materia / semana, cuatrimestre 2015-3

La LMS Chamilo permite la interacción entre alumnos, que debe ser promovida por el docente a fin de enriquecer el proceso educativo, esto a través de la instrucción expresa de formar equipos virtuales, observamos que (ver, Gráfico 4) el 86% no tuvo ninguna actividad que requiriera la interacción entre alumnos, el 10% generó una actividad que promovía la interacción entre alumnos y solo el 4% generó dos actividades que promovían la interacción de alumnos en plataforma.

Interacción entre alumnos

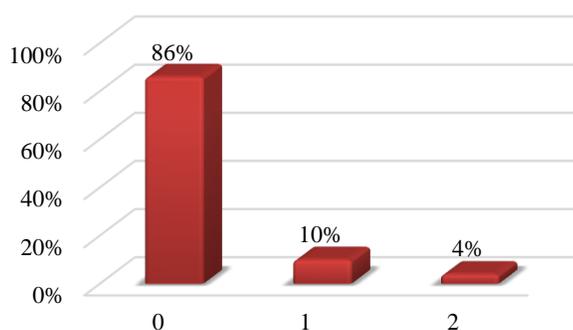


Gráfico 4 Promedio de número de interacciones entre alumnos a través de la generación de equipos de trabajos virtuales por cuatrimestre, cuatrimestre 2015-3

El número de herramientas de la LMS Chamilo que uso en promedio cada docente, durante el cuatrimestre 2015-3 (ver, Gráfico 5), fue como sigue: el 76% uso de entre 1 a 3 herramientas, el 11% uso de 4 a 6 herramientas, el 9% de 7 a 9 herramientas y solo el 2% más de nueve herramientas del total de herramientas disponibles en la LMS.

Herramientas utilizadas en plataforma

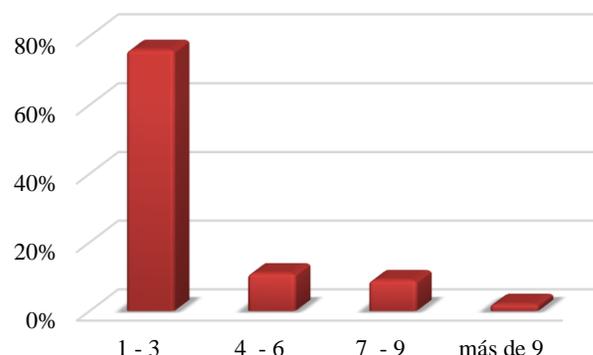


Gráfico 5 Promedio de número de herramientas de la LMS Chamilo, usadas por profesor, durante el cuatrimestre 2015-3

En la evaluación cualitativa, se registró lo siguiente, (ver, Tabla 1), existen tres herramientas más usadas por la muestra de docentes, estas son anuncios que es usada por el 93% de la muestra, tareas usada por el 86% y wikis usada por el 71%. Mientras que las herramientas absolutamente desaprovechadas fueron: evaluaciones, ejercicios y grupos.

Herramienta	Porcentaje
Anuncios	93%
Tareas	86%
Wikis	71%
Agenda	52%
Compartir documentos	51%

Tabla 1 Herramientas de la LMS Chamilo más usadas por la muestra de docentes

En el rubro de capacitación, durante el cuatrimestre 2015-3 se realizaron cursos sobre Competencias digitales para docentes y premisas de la educación b-learning (mismas que han sido objeto de otras investigaciones) para el manejo de la LMS, como acciones de profesionalización (ver Gráfico 6), el 10% acudió un promedio de 1 a 3 horas, el 30% de 4 a 6 horas, el 20% de 7 a 9 y el 40% más de 9 horas.

Acciones de profesionalización

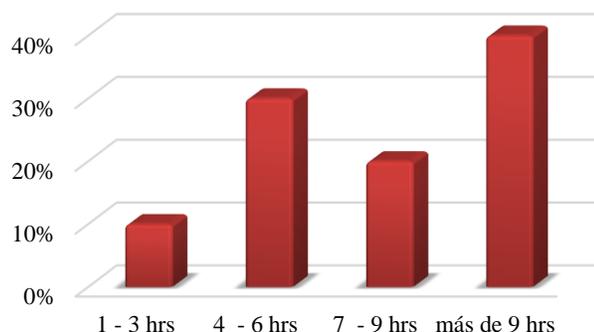


Gráfico 6 Número de horas invertidas por docente a acciones de profesionalización durante el cuatrimestre 2015-3

Evaluación de los contenidos.

De las cuatro actividades seleccionadas se omitirá el nombre de la actividad, así como la materia de la que fueron extraídas debido a que por la cantidad tan reducida de docentes que imparten clases en la modalidad, existe de facto la especialización en alguna materia, por lo cual se denominarán con las letras A, B, C y D para proteger la identidad de los docentes.

El ejercicio A corresponde a una materia del módulo directivo, el resultado del arbitraje de otros docentes-expertos (tres) según la rúbrica de evaluación fue como sigue:

Actividad "A" Materia del módulo directivo	Categorías de calificación				
	Muy adecuado	Adecuado	Regular	Inadecuado	Muy inadecuado
Correspondencia con los objetivos de la materia/unidad	60%	23%	17%		
Correspondencia de la exigencia con las competencias precurrentes que el alumno debe tener	55%	27%	18%		
Correspondencia entre la actividad y el tiempo indicado para hacerla	65%	35%			

Correspondencia entre la actividad y los instrumentos para evaluarla	45%	30%	15%	10%	
Correspondencia entre la actividad y el diseño instruccional	20%	10%	43%	27%	
Pertinencia de la actividad, dentro de la formación académica	90%	10%			
Percepción de que la actividad genere aprendizaje significativo	67%	20%	13%		
En general el diseño de la actividad fue	70%	25%	5%		
Incorporaría la actividad a su clase presencial, para aumentar la calidad educativa.	100%				

Tabla 2 Promedio de calificaciones obtenidas en la actividad A, en el arbitraje de docentes-expertos

En la tabla 2, podemos observar que en el arbitraje de los docentes-expertos la actividad fue adecuada, consideran tiene consistencia entre los objetivos de la materia y los objetivos de la actividad, el área con peores opiniones fue el diseño instruccional de la actividad.

Actividad "B" Materia Especializada	Categorías de calificación				
	Muy adecuado	Adecuado	Regular	Inadecuado	Muy inadecuado
Correspondencia con los objetivos de la materia/unidad	55%	45%			
Correspondencia de la exigencia con las competencias precurrentes que el alumno debe tener	45%	30%	5%	20%	
Correspondencia entre la actividad y el tiempo indicado para hacerla	100%				
Correspondencia entre la actividad y los instrumentos para evaluarla	70%	30%			
Correspondencia entre la actividad y el diseño instruccional	60%	20%	10%		
Pertinencia de la actividad, dentro de la formación académica	100%				
Percepción de que la actividad genere aprendizaje	100%				

significativo					
En general el diseño de la actividad fue	90%	10%			
Incorporaría la actividad a su clase presencial, para aumentar la calidad educativa.	80%	20%			

Tabla 3 Promedio de calificaciones obtenidas en la actividad B, en el arbitraje de docentes-expertos

En la tabla 3, observamos que las áreas mejor evaluadas fueron la pertinencia de la actividad y la percepción sobre la generación de aprendizajes significativos el área con menor porcentaje de evaluación fue la correspondencia de las competencias precurrentes del alumno y las exigencias de la actividad, aunque en las categorías Muy adecuado y adecuado suman más del 70%.

	Categorías de calificación				
	Muy adecuado	Adecuado	Regular	Inadecuado	Muy inadecuado
Actividad "C" Materia Especializada					
Correspondencia con los objetivos de la materia/unidad	100%				
Correspondencia de la exigencia con las competencias precurrentes que el alumno debe tener		89%	11%		
Correspondencia entre la actividad y el tiempo indicado para hacerla	100%				
Correspondencia entre la actividad y los instrumentos para evaluarla	100%				
Correspondencia entre la actividad y el diseño instruccional	90%	10%			
Pertinencia de la actividad, dentro de la formación académica	100%				
Percepción de que la actividad genere aprendizaje significativo	89%	11%			
En general el diseño de la actividad fue	100%				
Incorporaría la actividad a su clase presencial, para aumentar la calidad educativa.	100%				

Tabla 4 Promedio de calificaciones obtenidas en la actividad C, en el arbitraje de docentes-expertos

En la tabla 4, observamos que la actividad fue calificado en las categorías muy adecuado y adecuado excepto en las competencias necesarias para la ejecución de la actividad.

	Categorías de calificación				
	Muy adecuado	Adecuado	Regular	Inadecuado	Muy inadecuado
Actividad "D" Materia Especializada					
Correspondencia con los objetivos de la materia/unidad			100%		
Correspondencia de la exigencia con las competencias precurrentes que el alumno debe tener	100%				
Correspondencia entre la actividad y el tiempo indicado para hacerla	90%	10%			
Correspondencia entre la actividad y los instrumentos para evaluarla	80%	20%			
Correspondencia entre la actividad y el diseño instruccional	90%	10%			
Pertinencia de la actividad, dentro de la formación académica	100%				
Percepción de que la actividad genere aprendizaje significativo	100%				
En general el diseño de la actividad fue	100%				
Incorporaría la actividad a su clase presencial, para aumentar la calidad educativa.	100%				

Tabla 5 Promedio de calificaciones obtenidas en la actividad D, en el arbitraje de docentes-expertos

Por los porcentajes registrados de las evaluaciones observamos que son superiores al setenta por ciento (entre las categorías muy adecuado y adecuado) en cuanto a la correspondencia de la actividad con los objetivos de la materia y/o unidades temáticas, así como las competencias precurrentes que por el nivel formativo se deduce tiene el alumno y la pertinencia de la actividad. Sin embargo observamos que la correspondencia entre la actividad y el diseño instruccional es una de las

áreas con porcentajes más bajos en la evaluación.

Finalmente el total de los docentes-expertos incorporarían esta actividad a su clase presencial. En la sección para comentarios, se categorizaron en tres principales: la primera categoría, instrucciones a percepción de los docentes-expertos, varias de las instrucciones de las actividades son ambiguas o difíciles de comprender.

La segunda categoría tiempo suponen que el tiempo para la actividad es suficiente, sin embargo se cuestionaron cuanto tiempo se le exige al alumno dedique a la semana a su formación y la tercera categoría fue digitalización de contenidos, suponen que incorporar proyectos de esta naturaleza a la formación presencial, completaría la impartición de contenidos, aumentaría las horas de práctica y elevaría la calidad de la educación.

Conclusiones

La evaluación docente cuando es concebida como un medio necesario para la mejora de la educación y que cuyos resultados deben de estar vinculados con la evaluación del resto de las áreas y los actores en la educación ofrece la posibilidad de crecimiento y la colaboración de los docentes, pues no hay en ésta una creencia punitiva sino por el contrario la construcción de acciones que permiten profesionalizar al docente y dotarlo de las competencias necesarias para la ejecución exitosa de su labor.

Si bien una de las premisas que se ha planteado en la introducción obedece a realizar una evaluación sistémica, esta investigación no se ha realizado en aislado sino que forma parte de una serie de investigaciones que realiza el equipo cuya meta ambiciosa es lograr a corto plazo, la evaluación de la modalidad de ingenierías ejecutivas y colateralmente lograr la certificación del programa.

Entre los resultados más relevantes podemos resaltar el escaso número de herramientas de la LMS Chamilo que los profesores usan de manera regular, esto puede ser atribuible a la poca capacitación que estos recibieron al incorporarse al modelo semipresencial. De ahí que sea necesario profesionalizar la labor docente de las ingenierías ejecutivas, pues si bien se presupone una labor docente que se conoce tradicionalmente, también es necesario distinguir de manera precisa que las competencias necesarias si muestran diferencias pues son modos de ejecución distintas de la misma labor. Por ello la idea de la capacitar es necesaria y este diseño de actividades debe corresponder a los hallazgos en la evaluación, donde concluimos que el actuar en la LMS puede no ser el esperado, porque el docente no se encuentra familiarizado con ésta, de manera que le permita una interacción cotidiana que fomente el aprendizaje significativo.

Cualitativamente la evaluación de las actividades nos permite apuntar que el docente cuenta con los conocimientos teóricos necesarios para desarrollar un b-project, misma calidad que es reconocida por otros docentes al evaluar la actividad, sin embargo es necesario recalcar que la deficiencia en diseño instruccional nos dirige a la necesidad de dotar al docente de un bagaje de buenas prácticas en el desarrollo de contenidos en una LMS, incluso visualizado que varios de ellos puedan ser usados por docentes de la modalidad escolarizada. Finalmente, consideramos que es necesario resignificar las evaluaciones docentes de una acción punitiva y lastimosa para la integridad personal y profesional del docente, a una acción natural, objetiva y que promueva el desarrollo integral de los docentes.

Referencias

Area Moreira Manuel (2009) Introducción a la Tecnología Educativa: Manual electrónico. España: Universidad de la Laguna

Belloch, C. Teleformación. En Las TICs en las diferentes modalidades de enseñanza/aprendizaje. Universidad de Tecnología Educativa (UTE), Universidad de Valencia. <http://www.uv.es/belloch/pedagogia/EVA2.pdf> consultado: 12/05/16

Canales, A. (2004). Avances y retos de la evaluación de la docencia universitaria. En: Rueda Coordinador (2004). ¿Es posible evaluar la docencia en la universidad? México: ANUIES, UABJO. Disponible en: <http://prometeo3.us.es/publico/images/competencias.pdf> consultado: 16/04/16

Farley, Luis (2007) “Campus Virtual: la Educación más allá del LMS”, RUSC, vol. 4, n.1. disponible en: <http://www.uoc.edu/rusc/4/1/dt/esp/ortiz.pdf> consultado 24/05/16

González, V.M. Hernández, Z.G. (2011) Mododle para directivos. Guía para proyectos con ambientes virtuales de aprendizaje. México: UNAM

Loredo, J. (2004) “Principios orientadores para valorar la práctica docente. Un enfoque holístico”, en Rueda (coord.) (2004). ¿Es posible evaluar la docencia en la universidad? México: ANUIES, UABJO.

Marcelo Carlos. “Estudio sobre competencias profesionales para e-Learning”, disponible en:

Martinet, Raymond y Gauthier (2001) Estudio sobre competencias profesionales para e-learning. Junta de Andalucía - Proyecto Prometeo. Disponible en: <http://prometeo3.us.es/publico/images/competencias.pdf> consultado: 20/05/16

Monroy, M, Contreras, O. Desatnik, O. (2009) Psicología educativa. México: UNAM

Monroy, M. (2003) “Representaciones pedagógicas de estudiantes del nivel licenciatura sobre los profesores”. Tesis para obtener el grado de Doctor en Pedagogía. México: UNAM, Facultad de Filosofía y Letras.

Nó, Javier. (2005). “Planificación de la enseñanza semipresencial: una experiencia de posgrado”. Revista de Universidad y Sociedad del Conocimiento. 2005.Vol.2, No.1.

Valdés, H. (2007)” Evaluación del desempeño docente”, disponible en: <http://www.cei.es/de/rifad01.htm> consultado: 20/05/16

Valenzuela, Pineda Amparo Elizabeth. (2006) “Apuntes para una Educación Semipresencial” Guatemala: Universidad Rafael Landívar

Sistema de programación dinámica de trayectorias para asistencia de nodos bajo demanda en redes inalámbricas de sensores

GONZÁLEZ-SILVA, Marco Antonio*†, FRANCO-MORENO, Juan José, BAEZA-REGALADO, Michelle y MORENO-ZAMORA, Brenda Paola

Universidad Politécnica Metropolitana de Hidalgo

Recibido Noviembre 25, 2015; Aceptado Abril 01, 2016

Resumen

Proponer un sistema dinámico capaz programar recorridos eficientes donde se asista a los nodos de una red de sensores según las necesidades actuales de éstos. Lógico-deductiva, se parte de análisis de variables, se utilizan como base algoritmos de trazado de rutas, abstracción y uso de pruebas en casos particulares. Existen varios sistemas informáticos que emplean el uso de sensores para monitoreo y sensado de variables que son medibles en objetos o espacios, ejemplo de ellas son la temperatura de una habitación, la cantidad de humedad de un cultivo, el nivel de agua de un contenedor, la presión arterial de una persona, entre otros. En redes de sensores es común la planeación de trayectorias para visitar a los nodos con un determinado fin. Sin embargo, muchos de los algoritmos de trazado de rutas no toman en cuenta la prioridad y demanda actual de cada nodo en la red. En este artículo se presenta un sistema de programación de trayectorias para visitar nodos de una red según los valores que se obtengan de sus sensores, creando niveles de prioridad en cada uno de ellos. El sistema aquí presentado permite optimizar tiempos y recursos en la asistencia de nodos que requieran ser visitados según sus necesidades, aprendiendo de sí mismo para aumentar la precisión. Pruebas realizadas muestran como el sistema propuesto encuentra trayectorias óptimas para visitar nodos según demanda y prioridad. Estos resultados permiten considerar la creación de varios tipos de aplicaciones donde se optimicen recursos en el recorrido de una red.

Demanda, IoE, Sensores, wsn, Trayectorias

Citación: GONZÁLEZ-SILVA, Marco Antonio, FRANCO-MORENO, Juan José, BAEZA-REGALADO, Michelle y MORENO-ZAMORA, Brenda Paola. Sistema de programación dinámica de trayectorias para asistencia de nodos bajo demanda en redes inalámbricas de sensores. Revista de Análisis Cuantitativo y Estadístico. 2016. 3-7: 45-52

Abstract

There are several computer systems that make use of sensors in order to be able to monitor and sensing variables that can be measured in objects or spaces, precedent of them are the temperature of a room, the moisture of a crop, the water level of a container, blood pressure of a person, among others. In sensor networks is common the trajectory planning to visit nodes with a particular purpose. However, many of the traceroute algorithms do not take into account the priority and demand of each node in the network. This article presents a path programming system to visit nodes in a network according to the values obtained from the sensors, creating priority levels in each of them. The system here presented allows to optimize time and resources in support of nodes that require to be visited according to their needs, learning from itself to increase further accuracy. Tests show how the proposed system finds optimal paths to visit nodes on demand and priority. These results allow to consider the creation of various types of applications where it is required to optimize resources in the path of a network.

IoE, On-demand, Sensors, Trayectorias, wsn

*Correspondencia al Autor (Correo Electrónico: maagonzalez@upmh.edu.mx)

† Investigador contribuyendo como primer autor.

Introducción

El concepto de Internet de las cosas (IoE, por sus siglas en inglés Internet of Everything) se refiere al hecho de la interacción entre objetos, personas y máquinas conectadas a Internet. Su objetivo es el uso de tecnología como sensores, controladores y actuadores para el desarrollo de soluciones que ayuden a resolver distintos tipos de problemas, muchos de ellos de la vida cotidiana.

Charith, Chi y Srimal (2015), realizaron un estudio de distintos tipos de sistemas y aplicaciones desarrollados con IoE que tienen un alto impacto de utilidad en diferentes categorías como inteligencia portátil, casas, empresas y ciudades inteligentes (smart cities), motivando a desarrolladores a realizar trabajos sobre estos temas. Un ejemplo de ciudades inteligentes se puede ver en la propuesta de Asensio, Trasviña-Moreno, Blasco, Marco y Casas (2015) donde por medio de sensores se monitorea el nivel de tráfico de una avenida para programar cambios automáticos en los semáforos y agilizar el tránsito.

Como parte importante de IoE están las redes inalámbricas de sensores (WSN, por sus siglas en inglés Wireless Sensor Networks). Una WSN es una colección de sensores capaces de medir diferentes variables físicas como temperatura, ritmo cardíaco, nivel de líquido en un contenedor, humedad, aceleración, dirección de movimiento, presión de aire, etc., y por medio del uso de una señal de radiofrecuencia formar una red inalámbrica donde se puedan enviar datos hacia algún punto para ser procesados y/o analizados. De esta manera, acceder a datos sensados de algún objeto o persona puede realizarse desde diversos sitios remotos, utilizando la red de Internet como una opción y ampliando el panorama de aplicaciones al extender la cobertura de conectividad de las cosas y personas.

De acuerdo con Shane (2013), en un futuro una persona podrá tomar una píldora o tener sensores que detectarán su estado de salud e informar éste a través de Internet.

Muchas de las aplicaciones donde se hace uso de una WSN requieren que los nodos que forman parte de la red sean visitados con cierta periodicidad y propósito. Shue y Conrad (2013) presentaron un estudio del uso de robots como asistentes en las redes de sensores que visitan a los nodos con varios fines como localización, recolección de datos, reemplazar o reparar nodos dañados, realizar carga de baterías, entre otros. En un enfoque de aplicación con IoE, Vargheese y Dahir (2014) proponen el uso de sensores y análisis de datos para predecir información de productos fuera de stock en tiendas, y notificar a los asociados de ellas para programar visitas de sus proveedores para abastecimiento, de esta manera se prevé que los clientes siempre encuentren el producto buscado.

El problema de planear rutas para visitar diversos lugares donde se optimicen tiempos y recorridos se conoce como VRP (por sus siglas en inglés, Vehicular Routing Problem) y se planteó por primera vez por Dantzig y Ramser (1959). En aquél trabajo se presentó una solución para programar trayectorias de vehículos y satisfacer demandas de abastecimiento de gasolina en estaciones de servicio. A raíz de esto muchas variantes de VRP han aparecido en la literatura, por ejemplo se han agregado restricciones de tiempo para abastecer demandas en determinados puntos (VRP with Time Windows, conocido en inglés), y considerarse ésto en la planeación de trayectorias (Oliveira, Vasconcelos, Alvarenga, Mesquita y Souza, 2007). De manera similar, uno de los algoritmos más utilizados para programar recorridos de manera óptima y visitar nodos de una red conexa y no dirigida es el de caminos mínimos (Dijkstra, 1959).

Este algoritmo utiliza pesos en cada camino y explora todas las posibles rutas desde un origen hacia un destino buscando el menor costo según la suma de los pesos totales. Finalmente vale la pena citar al problema del agente viajero (TSP, en inglés Travelling Salesman Problem) citado en teoría de grafos y expuesto por W. Hamilton en el mismo año que Dijkstra.

TSP plantea el caso de encontrar la ruta óptima para recorrer un grafo donde se visiten todos los nodos y regresar al punto de partida, muchas soluciones computacionales se han expuesto para solucionar este problema (Reinelt G, 1994).

En general, existen varios sistemas y algoritmos para planeación de trayectorias propuestos en gran parte en áreas como teoría de toma de decisiones y puestos en práctica en redes de sensores. Sin embargo, estos algoritmos basan su planeación en estudios previos de demanda con datos no cambiantes y no toman en cuenta variaciones en los datos que pueden ser medidas en tiempo real y que afecten los requerimientos de asistencia en diferentes instantes de tiempo.

En este artículo se presenta un sistema llamado DynaSTy (por sus siglas en inglés Dynamic Schedule of Trayectories) para la programación dinámica de rutas y asistir a nodos que requieran ser visitados según los valores de sus variables medidas en tiempo real.

DynaSTy se basa en crear una red de sensores donde se mida algún parámetro que ofrezca información sobre una demanda existente y que interactúe con el entorno de un nodo cuyo estado sea dinámico respecto a ciertas entradas y salidas. Estos datos son enviados a través de una conexión inalámbrica hacia un servicio de base de datos en línea (cloud computing) donde podrán ser consultados en diversos instantes.

A su vez, un servidor local hace uso de la información en línea para programar rutas de asistencia mediante un algoritmo cuyos fundamentos son TSP y VRP donde las ponderaciones estarán en los nodos y no en los trayectos. El monitoreo en tiempo real permite al sistema crear curvas de aprendizaje útiles para predictibilidad, control y alcance, previniendo que los nodos sean asistidos según su demanda.

El contenido de este artículo se distribuye en cuatro secciones, en la sección 1 se mostró la introducción, en la sección 2 llamada metodología se presentaran los tres subsistemas de DynaSTy: red inalámbrica de sensores; cómputo en la nube y algoritmo de trazado de ruta. En la sección 3 se presentan pruebas y resultados de DynaSTy en simulador, y finalmente en la sección 4 se presentan las conclusiones.

Metodología

Como se mencionó anteriormente DynaSTy requiere de tres subsistemas para su funcionamiento los cuales se mencionan a continuación.

Subsistema de red inalámbrica de sensores

Una de las bases de DynaSTy es que los nodos que requieran ser asistidos deben sensar alguna variable física de su entorno y enviar los datos de ella a través de una conexión inalámbrica hacia una base de datos, donde serán almacenados y con base en ella crear una demanda de asistencia.

Utilizando una conexión por medio de tecnología 802.11 un sensor debe conectarse a la red de Internet ya sea de manera independiente o con ayuda de otro nodo.

En la segunda situación un sensor puede no tener la cobertura o capacidad de conectarse a Internet, pero deberá tener la posibilidad de enviar sus datos hacia otro nodo intermedio que a su vez se conecte a otro nodo intermedio y así sucesivamente hasta alcanzar una estación base o nodo sink, como se les conoce en WSN. Un nodo sink es capaz de recibir datos de otros nodos y tener conexión a Internet. En la figura 1 se aprecia como los datos de sensado del nodo A tienen que recorrer una trayectoria pasando por los nodos B, C (intermedios) y D donde éste último envía la información recibida hacia algún sitio en Internet. De esta manera se requiere que en el diseño de la red de sensores que participen exista una conexión a Internet de forma independiente o de manera asistida.

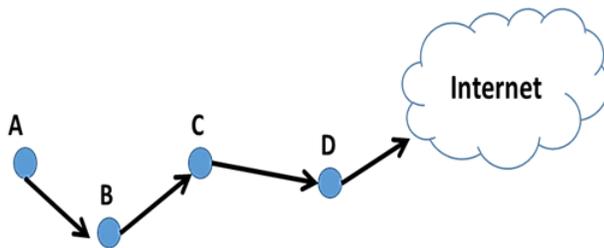


Figura 1 Conexión asistida hacia Internet en una red de sensores

Cómputo en la nube

Una vez que los sensores han enviado su información de sensado a una base de datos destinada, éstos representarán un valor de demanda (VDemanda). Esto es, si por ejemplo un sensor mide que la capacidad de un contenedor de líquido ha alcanzado el 70% de su capacidad éste tendrá un valor de demanda mayor que uno que tenga un 60%, tomando en cuenta que la demanda es recolectar el líquido. Para fines de DynaSTy, VDemanda será calculado siempre en porcentaje.

Los demás valores que requiere DynaSTy para su operación son los siguientes:

A. Ponderación. Esto es la razón de cambio de una variable respecto a otra variable (para el caso de DynaSTy el tiempo). Se calcula con la fórmula de razón de cambio promedio de la ecuación (1).

$$P = |\Delta x| / \Delta t \quad (1)$$

Donde:

P = Variación promedio (ponderación).

Δx = Valor final de x – Valor inicial de x. (valor absoluto)

Δt = Valor final de t – Valor inicial de t. (variación del tiempo).

Nótese que Δx está dado en valor absoluto porque para el funcionamiento de DynaSTy solo es relevante la variación promedio en términos positivos. Para ejemplificar la importancia de esta fórmula y concepto, el cual se ocupará en la sección 2.3, se considera el siguiente caso. Si en un contenedor se sensa que en 2 días el líquido que almacena subió del 50% al 70%, entonces aplicando la fórmula 1 tenemos:

$$P = (|50\% - 70\%|) / (2 \text{ días} - 0 \text{ días}) = 10\% / \text{día} \quad (2)$$

Entonces el resultado de (2) se puede interpretar como que el líquido contenido en un contenedor sube una razón promedio de 10% al día, lo que dará un valor de ponderación al nodo cuando se decida si debe ser visitado antes que otro nodo.

B. Identificador único (IdNodo). Número único asignado a un nodo para diferenciarlo de los demás.

C. Ubicación (Ubic). Es la información de latitud (Norte, Sur) y longitud (Este y Oeste) donde se encuentra localizado el nodo.

D. Fecha de asistencia (Fecha). Indica en qué fecha el nodo fue asistido y su demanda fue atendida por última vez.

E. Asistencia (a). Este valor indica si el nodo fue asistido en su demanda, y cambiará de acuerdo a la fecha de asistencia.

F. Contenido (ContNodo). Es el valor real en litros, kilogramos o cualquier unidad de medida según la variable que se esté sensando en el nodo. A diferencia de VDemanda este valor será representado en la unidad de medida usada.

Como medida adicional se considera un valor que define la capacidad de asistencia (CapA) del nodo que realizará la trayectoria calculada por DynaSTy. Dicho valor ayudará a obtener un estimado de cuántos nodos podrán ser asistidos antes que la capacidad del nodo visitante sea cubierta.

Algoritmo de trazado de ruta

Después de que todos los datos han sido almacenados y algunos de ellos estimados en la base de datos, se requiere calcular la ruta que deberá seguir un nodo que asiste para cubrir la demanda solicitada en la red. Este cálculo se realiza mediante el algoritmo de trazado de ruta que se explica a continuación.

Una premisa del algoritmo es que es necesario especificar los valores de Δt de la fórmula (1). El usuario puede decidir qué variación en el tiempo desea utilizar y que afectará en el valor de "P". Tal vez su variable principal, como puede ser el tiempo en el que un nodo agota su batería no represente ningún cambio si se toma un Δt con diferencia de unas cuantas horas, pero habrá una demanda distinta con un factor de variación de varios días o semanas, lo que obviamente afectaría en la programación de una ruta.

Otro requisito del algoritmo es que es necesario establecer un rango $R[a,b]$ respecto a VDemanda que indique cuando un nodo debe ser asistido. Por ejemplo siguiendo el mismo caso de demanda de abastecimiento de batería en un nodo, un valor de VDemanda dentro del rango de $R[0\%,69\%]$ indicará que un nodo debe ser asistido, porque se considera que no se tiene energía suficiente. Por lo tanto este valor tendrá un sentido distinto según los valores y criterios usados.

Una vez establecido $R[a,b]$, se obtendrá el valor de VDemanda de cada nodo dentro de la red. Si el valor consultado está dentro de $R[a,b]$ entonces este nodo ingresará a un nuevo arreglo llamado ArrayAsistibles, que indica qué nodos deben ser asistidos. Posteriormente para cada nodo dentro de ArrayAsistibles se consulta su valor de "P", calculado previamente por (1), y se crean nuevos arreglos donde se agrupan nodos con el mismo valor de ponderación. De esta manera, aquellos nodos que estén dentro de un arreglo con un valor de P más alto estarán el array1[a] y serán los primeros en ser asistidos, después aquellos que hayan sido clasificados con la segunda ponderación más alta y que estarán en el array2[b] y así sucesivamente hasta arrayN[n]. Para saber el orden de asistencia que deberá seguirse en estos últimos arreglos se considera un nodo inicial (el más cercano a un punto de partida donde el nodo asistente comenzará su recorrido, $array1[1] \in array1[a]$), posteriormente se visitará un segundo nodo ($array1[2]$) cuyos valores de longitud y latitud sean los más cercanos al primer nodo, y así sucesivamente hasta agotar el primer arreglo. Para el resto de los nodos incluidos en otros arreglos, por ejemplo el array2[b], se verificará desde un inicio si el nodo que asiste tiene la capacidad para visitarlos (CapA) y cumplir la demanda de sus nodos, de ser así se tomarán en cuenta éstos nodos en la ruta con la misma ponderación que $array1[a]$, considerando a ambos arreglos como uno solo.

Sin embargo, de no tener capacidad suficiente para cubrir la demanda de nodos de segundos, terceros u otros arreglos, entonces se considera la opción de asignar un segundo nodo asistente o simplemente programar más trayectorias respetando el orden establecido.

El pseudocódigo del algoritmo de trazado de rutas en presentado a continuación.

Pseudocódigo de algoritmo de trazado de rutas

Entrada: Arreglo de nodos en la red "A[x]", valores "P" para cada nodo, rango R[a,b].

Salida: Arreglo ordenado de nodos a asistir.

Inicio

//Obtener el arreglo ArrayAsistibles

1. Asignar: cont1 <-- 1, cont2 <-- 1
2. Asignar: aux <-- P(A[cont1])
3. If (aux >a)and (aux <b) then
4. Asignar: ArrayAsistibles[cont2]<-- A[cont1]
5. Asignar: cont2 <-- cont2+1
- end if
6. Asignar: cont1 <-- cont1+1
7. Regresar al paso 2 hasta que cont1 > x

//Obtener arreglos array1[a], array2[b]...arrayN[n] con distintas ponderaciones.

8. Asignar N <-- 1

9.-Determinar cuantos valores de P distintos hay en ArrayAsistibles (var1)

10.-Obtener el valor más alto de P (ValorP) asignado a un nodo dentro de ArrayAsistibles según la variable N. Si N=1 entonces se busca el primer valor más alto, si N=2 se busca el segundo valor más alto y así sucesivamente.

ValorP <-- mayorP(N,ArrayAsistibles[cont2])

//La función "mayorP" encuentra el valor de P según N buscado dentro del arreglo ArrayAsistibles.

1.-Buscar nodos dentro de ArrayAsistibles con P=ValorP.

12.- Crear el arreglo arrayN con los nodos encontrados en el paso 11.

13.- Asignar N<-- N+1

14.- Regresar al paso 10 hasta que N>var1

//Obtener ruta

15. Elegir el nodo inicial "i" de array1[a] y visitarlo
16. Crear lista de nodos visitados
17. Si el nodo más cercano a "i" no se ha visitado entonces
18. Hacer: visitar nodo j
19. Asignar: i <-- j
20. Actualizar lista de nodos visitados
21. Regresar al paso17 hasta que todos los nodos sean visitados
- 22.- Repetir los pasos 15-21 para los arreglos restantes array2[b], array3[c]..... arrayN[n]

En la figura 2 se muestra un bosquejo general de DynaSTy, donde se pueden apreciar los tres subsistemas. En primera instancia nodos de una red sensan ciertas variables y envían datos hacia algún sitio en Internet el cual sirve de intermediario para la consulta que se hace desde un sitio remoto para alimentar al algoritmo que calcula la ruta a seguir para asistir a los nodos según su demanda. Los nodos en la figura presentan un valor numérico que indica su valor de demanda En el lado izquierdo se visualiza una ruta que recorrerá primeramente los nodos con demanda 1.

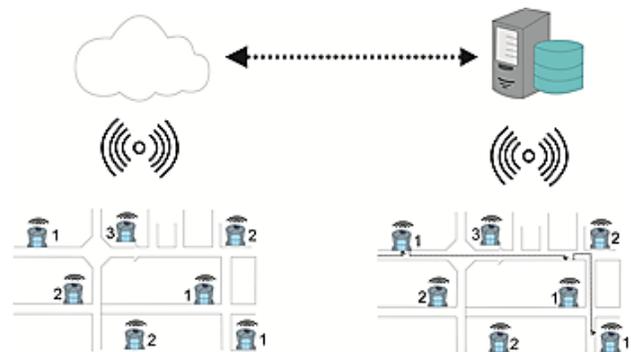


Figura 2 Bosquejo general de DynaSTy

Pruebas y resultados de DynasSTy

Para probar el desempeño de DynasSTy se programó éste en el software Matlab versión 7.6.0. con los siguientes valores.

Se creó un arreglo de 30 nodos como entrada A[30]. Se asignaron de manera aleatoria valores de VDemanda inicial y final a cada uno de los nodos y se crearon dos arreglos con estos valores B[30] y C[30]. Se estableció un rango R de [0,60]. Se definió una variación de tiempo $\Delta t = 2$ días. La capacidad del nodo asistente se consideró que solo puede asistir por día a 15 nodos. Se realizaron 3 iteraciones del algoritmo de trazado de rutas donde en cada iteración se crearon nuevos valores de B[30] y C[30] respetándose los resultados de la iteración anterior, es decir si el nodo había sido asistido anteriormente entonces su valor VDemanda inicial tendría 0%.

Adicionalmente se programó en Matlab recorridos utilizando el método del nodo más cercano, es decir, a partir de un nodo inicial, visitar su vecino más cercano y posteriormente el más cercano a este último y así sucesivamente hasta recorrer todos los nodos. Para este método se tomó en cuenta como inicio una versión del arreglo ArrayAsistibles que incluye todos los 30 nodos ordenados según su valor de VDemanda. Para siguientes iteraciones se contempló el mismo orden que la primera iteración y descartar así los cambios en demanda subsecuentes de los siguientes días. A este método le llamamos VCP (Vecinos Cercanos sin Ponderación).

En el Gráfico 1 se pueden apreciar los resultados de DynaSTy comparado con VCP. En el eje x se muestran los 30 nodos de las pruebas y en el eje y se muestra el valor promedio alcanzado por VDemanda de cada nodo durante los 6 días. La línea continua muestra el resultado de DynaSTy y la línea punteada de VCP. Mientras que en DynaSTy solo dos nodos estuvieron sin asistir cuando su valor de demanda bajo a menos del 70%, en VCP casi cerca de la mitad de los nodos su demanda no fue cubierta, sobre todo aquellos nodos al final del arreglo.

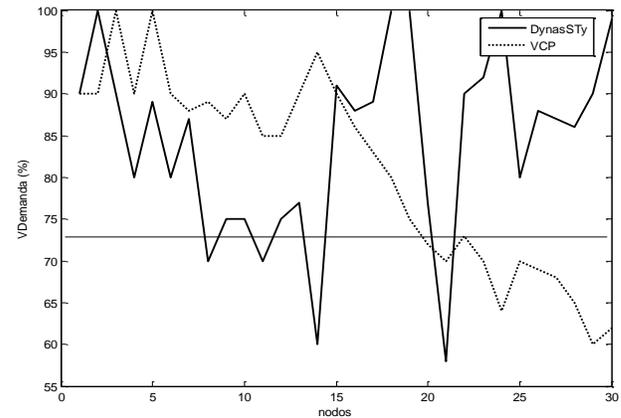


Gráfico 1 Comparativa entre DynaSTy y VCP

Lo que se puede concluir es que aquellos nodos que requerían asistencia con DynaSTy no fueron cubiertos en parte por una demanda muy alta generada aleatoriamente y en parte por la capacidad del nodo que asistía. En cambio en VCP además de estos factores influyó no tomar en cuenta la variabilidad y estancarse en rutas predefinidas con datos no cambiantes.

Conclusiones

En teoría de decisiones, redes de sensores e IoE, es importante considerar los cambios que presenten variables que influyen directamente en condiciones de entorno y que pueden modificar el comportamiento de un sistema completo.

La asistencia bajo demanda que se realice hacia nodos de una WSN puede prevenirse de manera que se eviten problemas como el que cierta sustancia se agote, que se pierda alguna conexión, que se contamine en exceso alguna región, o simplemente que los clientes de un establecimiento siempre encuentren lo que buscan.

En este artículo se presentó un sistema llamado DynaSTy que combina de redes de sensores, cobertura amplia de redes como Internet y cómputo que realiza cálculos donde se tomen en cuenta valores de variables medidas en tiempo real para asistir a nodos que tienen cierta demanda y que esta puede variar con el tiempo. DynaSTy puede tener la opción de ajustarse a periodos de tiempo para predecir situaciones de demanda a futuro o asistir a aquellas en tiempo real.

DynaSTy podría tener un mayor desempeño y otras funcionalidades si se toman en cuenta más de una variable que puedan afectar el entorno de un nodo.

Agradecimientos

Agradecemos el apoyo brindado de la Universidad Politécnica Metropolitana de Hidalgo.

Referencias

Asensio A., Trasviña-Moreno, Blasco R., Marco A. & Casas R. (2015). Wireless Sensor Network in traffic management system. International Conference on Applied Informatics and Computing Theory. Salerno, Italy, (Junio 2015), 60-68

Charith P., Chi Harold L. & Srimal J. (2015). The Emerging Internet of Things Marketplace From an Industrial Perspective: A Survey. IEEE Trans. Emerg. Top. Comput. 3, 4 (October 2015), 585-598.

Dantzig G. B. & Ramser J. (1959). The Truck Dispatching Problem. Management Science 6 (1): 80–91.

Dijkstra E.W. (1959). A Note on Two Problems in Connexion with Graphs. Numerische Mathematik, 1, 269-271.

Oliveira H.C., Vasconcelos G.C., Alvarenga G.B., Mesquita R.V. & Souza M.M. (2007). A robust method for VRPTW with Multi-Start simulated annealing and statistical analysis. IEEE Symposium on Computational Intelligence in Scheduling. Honolulu, HI, USA, (Abril 2007), 198-205

Reinelt G. (1994). The travelling Salesman: Computational Solutions for TSP Applications. Springer-Verlag, Berling.

Shane M., Villa N., Stewart-Weeks & Lange Anne. (2013) The Internet of Everything for Cities, Connecting People, Process, Data, and Things To Improve the “Livability” of Cities and Communities. Acceso en Mayo 2016, <http://www.cisco.com/web/strategy/docs/gov/everything-for-cities.pdf>

Shue S. & Conrad J.M. (2013). A Survey of Robotics Applications in Wireless Sensor Networks. Southeastcon. Jacksonville, Fla, USA, (Abril 2013). 1-5

Vargheese R. & Dahir H. (2014). An IoT/IoE enabled architecture framework for precision on shelf availability: Enhancing proactive shopper experience. IEEE International Conference on Big Data. Washington, DC, USA. (October 2014). 21-26

Un algoritmo transgénico vs algoritmos de optimización

CORREA-VILLALÓN, Christian José†, LUNA-RAMÍREZ, Enrique*, SORIA-CRUZ, Jorge y CRUZ-VALENZUELA, Roberto

Instituto de Educación de Aguascalientes. San Luis Potosí 601, Ojocaliente. Aguascalientes, Ags. C.P. 20256
Instituto Tecnológico El Llano Aguascalientes. Km. 18 Carr. Ags.-S.L.P. El Llano Aguascalientes. C.P. 20330

Recibido Octubre 29, 2015; Aceptado Marzo 12, 2016

Resumen

En este artículo se presenta un algoritmo transgénico diseñado para resolver problemas de optimización basándose en el tratamiento de funciones engañosas. Para ello, se incorporan elementos novedosos al algoritmo transgénico: homología de genes, alineamiento de cromosomas, resolución positiva-negativa y replicación de cultivos. De esta manera, es posible encontrar los individuos más aptos de una población, lo cual es discutido mediante la presentación de los resultados obtenidos de la aplicación del algoritmo a un grupo de funciones engañosas, cuya finalidad es precisamente demeritar la eficacia de las técnicas heurísticas. Tales resultados fueron comparados con resultados de otros algoritmos de optimización en el estado del arte, haciéndose evidente la eficiencia de nuestro algoritmo y concluyéndose que éste no requiere de poblaciones grandes, ni de un gran número de generaciones para poder obtener buenos resultados.

Algoritmos transgénicos, Problemas de optimización

Abstract

In this paper, a transgenic algorithm designed to solve optimization problems, based on the treatment of deceptive functions, is presented. To deal with it, innovative elements are incorporated to the transgenic algorithm: homology of genes, alignment of chromosomes, positive-negative resolution and crop replication. In this way, it is possible to find the best individuals of a population, as discussed through the presentation of the results obtained from the algorithm application to a group of deceptive functions, which purpose is just to break the effectiveness of heuristic techniques. Such results were compared with the results of other optimization algorithms in the state of art, which allowed to show the efficiency of our algorithm, concluding that it does not require big populations, neither a big number of generations in order to obtain good results.

Transgenic algorithms, Optimization problems

Citación: CORREA-VILLALÓN, Christian José, LUNA-RAMÍREZ, Enrique, SORIA-CRUZ, Jorge y CRUZ-VALENZUELA, Robert. Un algoritmo transgénico vs algoritmos de optimización. Revista de Análisis Cuantitativo y Estadístico. 2016. 3-7: 53-61

*Correspondencia al Autor (Correo Electrónico: elunaram@hotmail.com)

† Investigador contribuyendo como primer autor.

Introducción

La inspiración para la creación de la heurística propuesta en este artículo es la Biotecnología, sobre la que existen diversas definiciones, siendo una de las más relevantes la definición mencionada en la Convention on Biological Diversity (1993), abordada ampliamente por Celesti et al. (2016) en su tratado sobre sistemas computacionales aplicados al procesamiento masivo de secuencias de DNA. Esta definición enuncia que “La Biotecnología es cualquier aplicación tecnológica que haga uso de sistemas biológicos, organismos vivos o derivados, para crear o modificar productos o procesos para un uso específico”. Un ejemplo típico de esto es la manipulación genética de organismos vivos, razón por la cual se les conoce como organismos transgénicos (Villalobos, 2007). Es importante mencionar que los cambios aplicados a un organismo transgénico se realizan buscando un objetivo específico, relacionado con la mejora de algunas de sus características.

En la actualidad existen diversos ejemplos de organismos transgénicos, que pueden ir desde modificaciones genéticas en cosas tan comunes como los cereales que consumimos a diario, hasta la modificación genética de una planta de tabaco, a la cual se le puede agregar el gen que permite generar luz a las luciérnagas, de manera que al entrar en contacto con una sustancia específica, la planta pueda emitir una radiación de luz (Ow et al., 1986). Así, el primer algoritmo transgénico (Correa et al., 2007) fue diseñado con base en el proceso de mejoramiento genético en las plantas. De esta manera, en un primer intento por emular este proceso, sólo fueron usadas algunas bases teóricas. No obstante, en la siguiente sección, se describe una versión mejorada del algoritmo transgénico, que considera factores significativos en la producción de transgénicos, que no habían sido considerados en trabajos anteriores.

Algoritmo Transgénico con Homología

Apoyándose en el uso de técnicas para la creación de individuos modificados genéticamente (Ignacimuthu, 2008), se diseñó una nueva versión del algoritmo transgénico, mostrada en la figura 1. Como se puede observar en la figura, este algoritmo incorpora nuevas características tales como la selección elitista inversa, que, a diferencia de la selección elitista, permite determinar los individuos menos aptos en una población, cuyos genes necesariamente desaparecerán en generaciones posteriores. De esta manera, las subpoblaciones conformadas por porcentajes de las poblaciones de los individuos más aptos y menos aptos, conocidas como población positiva y población negativa, respectivamente, son sometidas a la técnica de homología de genes, que consiste en un alineamiento de cromosomas mediante el cual es posible identificar el dominio de un gen sobre otros. Esta técnica es precisamente la que le da el nombre a nuestro algoritmo, Algoritmo Transgénico con Homología (ATH).

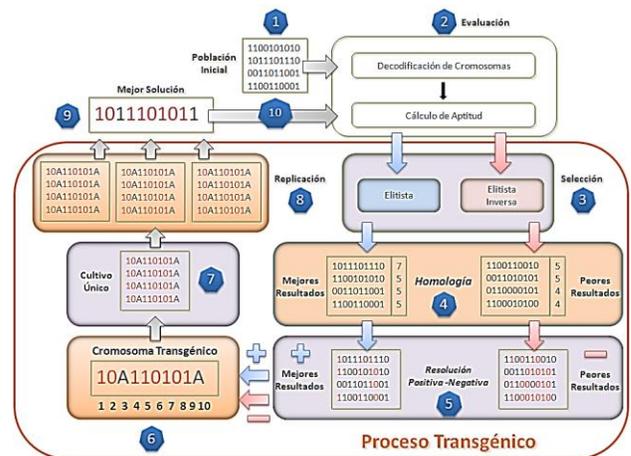


Figura 1 Algoritmo Transgénico con Homología

Antes de proceder a la descripción del ATH, es preciso resaltar otras características importantes del mismo, como es el caso de la combinación de las llamadas resolución positiva y resolución negativa, que al operar juntas, se generan mejores resultados, dado que en una misma generación, los genes que producen buenos resultados son preservados, mientras que los genes que producen malos resultados son rechazados (Correa et al., 2010).

La replicación de cultivos en una misma generación es otra característica importante del ATH, ya que permite llegar más rápidamente a la solución óptima, como se discutirá en la presentación de resultados.

Creación de la población inicial. La población inicial para comenzar a operar el ATH se crea en forma totalmente aleatoria, como sucede con otras técnicas heurísticas, por ejemplo, los algoritmos genéticos (Goldberg, 1989).

Así, en el paso 1 de la figura 2, se pueden observar soluciones iniciales conformadas por ceros y unos.

Evaluación. En esta etapa se evalúa la población inicial, obteniéndose para cada individuo el nivel de satisfacción de la función objetivo (la función que será optimizada). En nuestro caso, se optó por utilizar la función conocida como OneMax (Schaffer & Eshelman, 1991), dado que se trata de una función típica para medir la eficiencia de heurísticas, la cual básicamente mide la máxima cantidad de elementos "1" en cada individuo. Esto implica que los datos contenidos en cada individuo deben ser decodificados, de manera que sea posible determinar dicha cantidad. Hasta este punto, el ATH opera de manera similar a otras técnicas heurísticas; no obstante, en lo sucesivo se mostrará su parte original.

Selección. En el mundo de las técnicas heurísticas, existen diferentes tipos de selección de individuos. Para el caso del ATH, se optó por una selección elitista, cuyo objetivo es identificar a los individuos más aptos de una población. Así, si se tratara de un algoritmo genético, los individuos más aptos serían aquellos que mejor satisfacen la función objetivo, por lo que serían usados para realizar una cruce entre ellos y poder generar nuevos individuos con mejores características. En nuestro algoritmo, la selección elitista juega el mismo papel, excepto que los individuos no son cruzados entre sí, sino más bien son sometidos a un alineamiento de cromosomas, como se verá más adelante. Sin embargo, la principal diferencia respecto a otras heurísticas radica en lo que hemos denominado selección elitista inversa, cuyo objetivo es identificar a los individuos menos aptos, entendiéndose éstos como aquellos que se alejan de una posible solución.

Homología. Una vez identificadas las poblaciones de individuos más aptos e individuos menos aptos, se derivan de ellas subpoblaciones tomando un porcentaje de cada una, comúnmente entre el 30% y 40%, dependiendo del tamaño del problema (Correa et al., 2010). Estas subpoblaciones, conocidas como población positiva y población negativa, son sometidas a la técnica de homología de genes. En un plano epistemológico, homología se refiere a la correspondencia que existe entre dos o más rasgos de diferentes organismos, es decir, a la similitud que existe entre ellos. No es una relación directamente observable, sino más bien es una relación que se infiere a través de algún método de análisis.

Así, la homología es usada en el proceso de creación de organismos transgénicos para detectar patrones existentes en los genes, tanto de organismos receptores de nuevos genes, como de organismos donadores.

Con base en esto, es posible determinar qué genes conviene modificar y qué genes no deben de ser modificados en un organismo. En el contexto computacional, la homología también tiene el propósito de detectar patrones en los cromosomas de los individuos de una población, lo cual es realizado mediante un alineamiento de cromosomas, cuya idea básica es representar y comparar dos o más secuencias/cadenas de ADN u otras estructuras proteicas primarias con el objeto de resaltar sus zonas de similitud, mismas que podrían indicar relaciones funcionales o evolutivas entre genes (De Luna & Mishler, 1996). Bajo esta idea, los cromosomas (secuencias de ceros y unos) son almacenados como filas de una matriz, en los cuales se insertan espacios para que las zonas con estructuras idénticas o similares se alineen.

En la figura 2 se presenta un típico alineamiento de cromosomas de la población positiva, en la cual se realiza la búsqueda de patrones mediante el recorrido de cada una de las columnas que conforman esta población. Como se puede observar en la figura, solo fueron seleccionados cuatro de los diez individuos de la población, resaltándose sus genes dominantes con color rojo y encerrados en círculo, identificados con base en el mayor número de presencias (ceros o unos) en cada columna. Por ejemplo, la primera columna contiene un gen dominante correspondiente al número “1”, por lo que éste prevalecerá en el siguiente cultivo en esa misma posición. Por el contrario, un ejemplo donde no se observa un gen dominante es en la segunda columna, en la cual existe la misma cantidad de ceros y unos, por lo que se tendrá un valor aleatorio en esta posición en el siguiente cultivo. Este criterio se aplica por igual a las otras columnas donde no existe un gen dominante (columnas 3, 4, 6, 9 y 10).

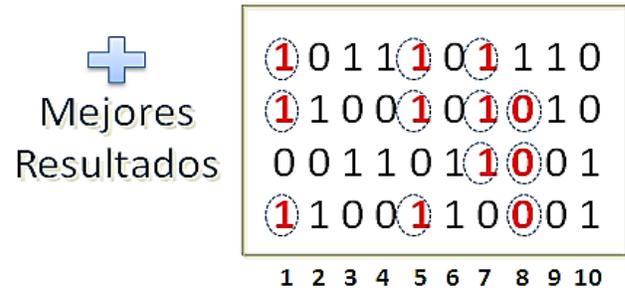


Figura 2 Alineamiento positivo de cromosomas

Al finalizar el alineamiento de cromosomas en la población positiva, se obtiene un cromosoma transgénico positivo o vector K, utilizado para crear el siguiente cultivo. En la figura 3 se muestra dicho cromosoma, el cual se mantiene el gen dominante “1” en las posiciones 1, 5, 7 y el gen dominante “0” en la posición 8.

Para las otras posiciones, como se mencionó, se generan valores aleatorios, indicados en la figura con la letra “A”.

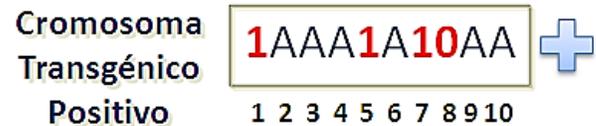


Figura 3 Cromosoma resultante del alineamiento

El alineamiento de cromosomas sobre una población negativa sigue el mismo procedimiento descrito anteriormente, por lo que el resultado también es un cromosoma transgénico, llamado en este caso cromosoma transgénico negativo o vector K' (véase figura 4). Sin embargo, contrario al caso del vector k, los genes dominantes de este cromosoma serán rechazados en la siguiente generación, debido a que éstos representan resultados no deseados al haber sido derivados de una población negativa.

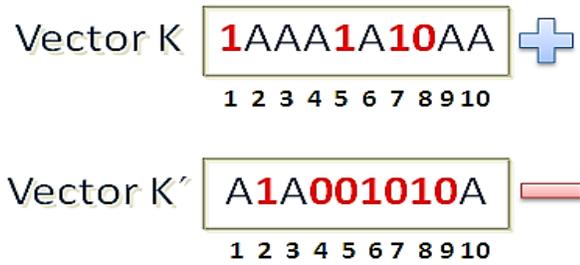


Figura 4 Cromosomas positivo y negativo juntos

Replicación. Una vez creado el cromosoma transgénico, se procede a generar el siguiente cultivo con base en este cromosoma. En esta etapa, el ATH echa mano del principio usado en Biotecnología para generar una población transgénica: para obtener mejores resultados a lo largo de nuevas generaciones, se puede utilizar un único cultivo o se pueden utilizar réplicas de un mismo cultivo. De esta manera, la mejor solución encontrada en una generación es incorporada a la siguiente generación, repitiéndose el algoritmo a partir del segundo paso hasta llegar a la mejor solución posible o, en su defecto, a la condición especificada para detener el algoritmo.

Resultados

En esta sección se discuten los resultados obtenidos de someter el ATH al tratamiento de funciones engañosas, cuya finalidad es precisamente demeritar la eficacia de técnicas heurísticas. De manera concreta, estas funciones permiten indagar la forma en qué los algoritmos orientados a la optimización fallan en la determinación de óptimos globales.

Las funciones engañosas están asociadas a lo se conoce como decepción, que se refiere al estado de reducción en la aptitud de un individuo generado por una combinación de buenos elementos constructores, cuando la lógica indicaría lo contrario, un aumento en la aptitud.

Este fenómeno fue sugerido originalmente por Goldberg (1989) para explicar el mal desempeño de los algoritmos genéticos en algunos problemas de optimización. Para comprender mejor esto, considérese el siguiente ejemplo relativo a una función de aptitud (función engañosa), cuyo dominio está compuesto por cadenas binarias de longitud 3 y su contradominio son propiamente los valores de aptitud:

Cadena	Aptitud	Cadena	Aptitud
000	70	100	30
001	50	101	2
010	49	110	3
011	1	111	80

Tabla 1

En este ejemplo se puede observar que la tendencia es a tener una mayor aptitud, a medida que una cadena contiene más ceros; no obstante, el óptimo global se localiza en la cadena que contiene sólo unos. En este caso, un algoritmo genético tendería típicamente a ir seleccionando cadenas con más ceros en la búsqueda de una solución óptima, por lo que en la cadena “111”, que es propiamente el óptimo global en este ejemplo, nunca sería determinada como tal. Con base en estas consideraciones, el ATH fue sometido a diversas funciones engañosas y los resultados obtenidos fueron comparados con los resultados de dos algoritmos trascendentales en el estado del arte, a saber, un algoritmo evolutivo que utiliza mapas de Markov para modelar soluciones prometedoras y un algoritmo genético con varios tipos de cruzamiento. En el Evolutionary Algorithm Based on a Markov Graphical Model Selection of Promising Solutions (Ponce et al., 2004), al cual se hará referencia como algoritmo evolutivo, los autores proponen una variante del algoritmo genético, que básicamente consiste en realizar la etapa de selección usando mapas de Markov.

Este algoritmo fue sometido a diversas funciones engañosas, en particular a la función conocida como FC₂ (Mühlenbein et al., 1998), considerada en la actualidad como una de las funciones con mayor grado de complejidad, razón por la cual fue seleccionada para realizar nuestro análisis comparativo entre algoritmos. La función FC₂ se define de la siguiente manera:

$$FC_2 = \sum_{j=1}^m F^5_{muhl}(s_j)$$

donde $s_j = (x_{5j-4}, x_{5j-3}, x_{5j-2}, x_{5j-1}, x_{5j})$ (1)

$$F^5_{muhl}(x) = \begin{cases} 3.0 & \text{for } x = (0,0,0,0,1) \\ 2.0 & \text{for } x = (0,0,0,1,1) \\ 1.0 & \text{for } x = (0,0,1,1,1) \\ 3.5 & \text{for } x = (1,1,1,1,1) \\ 4.0 & \text{for } x = (0,0,0,0,0) \\ 0.0 & \text{otherwise} \end{cases}$$

En la siguiente tabla se presentan los parámetros de experimentación, habiéndose utilizado los mismos valores para el tamaño de cadena y el tamaño de población para ambos algoritmos, esto con la finalidad de que operen en condiciones similares:

Algoritmo Evolutivo	
Problema	Valores para FC ₂
Tamaño de Cadena	10
Tamaño de Población	100,300
Porcentaje de Selección	70%
ATH	
Problema	Valores para FC ₂
Tamaño de Cadena	10
Tamaño de Población	100,300
% sel. positiva, negativa	2%, 2%
Porcentaje de dureza	90%

Tabla 2

Nótese que en la tabla anterior aparece el concepto de dureza, el cual se refiere al criterio (porcentaje) utilizado para determinar si un gen es dominante. Así, un 90% de dureza se refiere a que el número de veces de aparición de un elemento (“0” o “1”) en una columna debe ser al menos el 90% del total de elementos en esa columna para que pueda ser considerado como un gen dominante. Los parámetros anteriores produjeron los siguientes resultados:

Algoritmo Evolutivo		
Problema	FC ₂	FC ₂
Tamaño de Cadena	10	10
Tamaño de Población	100	300
Medida de generaciones	24	18
ATH		
Problema	FC ₂	FC ₂
Tamaño de Cadena	10	10
Tamaño de Población	100	300
Medida de generaciones	6	2

Tabla 3

En esta tabla, la media de generaciones es el resultado que permite medir la eficiencia de los algoritmos, la cual se refiere al número de generaciones que se requirieron para alcanzar el óptimo de la función FC₂. Así, los resultados muestran que el algoritmo ATH es significativamente más eficiente que el algoritmo evolutivo, ya que para ambos tamaños de población (100 y 300), se requirieron significativamente menos generaciones para alcanzar el óptimo.

Heywood (2006) muestra el comportamiento de un algoritmo genético ante un problema engañoso usando diferentes tipos de cruzamiento. El problema engañoso usado por Heywood, relativo a la teoría de bloques constructores, se muestra en la siguiente tabla:

$F(0^{**}) > f(1^{**})$	$F(00^*) > f(11^*), f(01^*), f(10^*)$
$F(*0^*) > f(*1^*)$	$F(0^*0) > f(1^*1), f(0^*1), f(1^*0)$
$F(**0) > f(**1)$	$F(*00) > f(*11), f(*01), f(*10)$

Tabla 4

Para el caso particular de orden 3, es decir, cuando se utilizan cadenas de 3 bits en el dominio de la función f , el problema asume los valores mostrados en la siguiente tabla:

$f(000)=28$	$f(001)=26$
$f(010)=22$	$f(100)=14$
$f(110)=0$	$f(011)=0$
$f(101)=0$	$f(111)=30$

Tabla 5

En la siguiente tabla se presentan los parámetros de experimentación, habiéndose incrementado a 30 el número de bits en las cadenas (Bloques Constructores con 10 subcadenas de Orden 3 – $BCO^{3(10)}$), aumentándose por tanto de manera significativa la complejidad del problema:

Algoritmo Genético	
Problema	$BCO^{3(10)}$
Tamaño de cadena	30
Tamaño de población	10,200,2000
Cruzamiento	1 Punto, Uniforme, "Tagged"
Evaluaciones	300,10000,50000
ATH	
Problema	Valores para FC_2
Tamaño de la cadena	30
Tamaño de la población	10,200,2000
%sel. Positiva, negativa	2%, 2%
Evaluaciones	300,10000,50000
Número de cultivos	2
Porcentaje de dureza	90%

Tabla 6

Los resultados obtenidos para ambos algoritmos, el genético y el ATH, fueron los siguientes:

Algoritmo Genético			
Problema	$BCO^{3(10)}$	$BCO^{3(10)}$	$BCO^{3(10)}$
Tamaño de la población	10	200	2000
Evaluaciones	300	10000	50000
Cruzamiento		1 Punto	1 Punto
Porc Resolución		27%	38%
Cruzamiento	No reporta	Uniforme	Uniforme
Porc Resolución		27%	35%
Cruzamiento		"Tagged"	"Tagged"
Porc Resolución		53%	64%
ATH			
Problema	$BCO^{3(10)}$	$BCO^{3(10)}$	$BCO^{3(10)}$
Tamaño de población	10	200	2000
Evaluaciones	300	10000	50000
Porc Resolución	60%	70%	100%

Tabla 7

Como lo muestran los resultados, también en este caso el ATH es más eficiente en la determinación de la solución óptima, al haber logrado siempre un mayor porcentaje de resolución del problema, llegando incluso a una resolución del 100% con un tamaño de población y un número de evaluaciones suficientes (2000 y 50000, respectivamente).

Conclusiones y trabajo futuro

Se presentó un algoritmo transgénico (ATH) diseñado para resolver problemas de optimización tomando como base el tratamiento de funciones engañosas, las cuales permiten indagar la forma en que los algoritmos orientados a la optimización fallan en la determinación de óptimos globales. El diseño de un primer algoritmo transgénico estuvo basado en el proceso de mejoramiento genético en las plantas, de manera que en un primer intento por emular este proceso, sólo fueron usadas algunas bases teóricas. No obstante, en este trabajo, se describió una versión ampliada y mejorada del algoritmo, que considera otros factores importantes en la producción de transgénicos, tales como homología de genes, alineamiento de cromosomas, resolución positiva-negativa y replicación de cultivos.

Nuestro algoritmo fue sometido a diversas funciones engañosas y los resultados obtenidos fueron comparados con los resultados de dos algoritmos recientes y de trascendencia en el estado del arte: un algoritmo evolutivo que utiliza mapas de Markov para modelar soluciones prometedoras y un algoritmo genético con varios tipos de cruzamiento. Un análisis comparativo mostró que el ATH es significativamente más eficiente que ambos algoritmos, ya que en la comparación con el algoritmo evolutivo, se observó que se requieren significativamente menos generaciones para llegar a la solución óptima, además de que la razón de eficiencia del ATH aumentó al aumentar el tamaño de la población inicial. Respecto a la comparación con el algoritmo genético, el ATH logró siempre un mayor porcentaje de resolución, llegando incluso a una resolución del 100% utilizando un tamaño de población y un número de evaluaciones suficientes, lo que al algoritmo genético nunca logró (su resolución máxima fue del 64%). Además, los resultados mostraron que el ATH no requiere de un tamaño de población grande para lograr buenos resultados. Como trabajo futuro, se tiene considerado realizar más análisis comparativos con algoritmos de optimización que operan bajo la filosofía de colonia de hormigas (Dorigo & Stützle, 2004), que cuenta con una aceptación importante en la actualidad debido a desempeño en problemas de optimización combinatoria.

Referencias

- Convention on Biological Diversity. Use of Terms. United Nations Treaty Series, Vol. 1760 I-30619. 1993.
- Celesti, A., Fazio, M., Celesti, F., Sannino, G., Campo, S., Villari, M. New Trends in Biotechnology: The Point on NGS Cloud Computing Solutions. En: IEEE Workshop on ICT solutions for eHealth 2016.
- Correa, C. J., Padilla, F., Pérez, E. & Quezada, F. Algoritmo Transgénico para resolver el problema del TSP. En: Tercer Congreso Mexicano de Computación Evolutiva COMCEV' 07. Aguascalientes, México, 2007, pp. 75-78.
- Correa, C. J., Padilla, F., Pérez, E., Ponce de León, E., Quezada, F., Ponce, J. & Padilla, A. New Algorithm Transgenic with Homology to Solve the Problem of OneMax. Revista Computación y Sistemas del CIC-IPN, 13(4): 3-8. Abril-Junio 2010.
- De Luna, E. & Mishler, B.D. El concepto de homología filogenética y la selección de caracteres taxonómicos. Boletín de la Sociedad Botánica de México, 59(1): 131-146. Mayo 1996.
- Dorigo, M. & Stützle, T. Ant Colony Optimization. MIT Press. 2004.
- Goldberg, D.E. Genetic Algorithms in Search, Optimization and Machine Learning. Addison-Wesley. 1989.
- Heywood, M.I. Deception in Genetic Search. <http://web.cs.dal.ca/~mheywood/CSCI6506/HandOuts/N04-Deception.pdf>. 2006. Sitio visitado el 31 de agosto de 2016.
- Ignacimuthu, S. Biotechnology: An Introduction. Alpha Science International, Ltd. 2008.
- Mühlenbein, Mahnig, and Rodriguez O. Schemata, distributions and graphical models in evolutionary optimization. Journal of Heuristics, 5: 215-247. 1998.

Ow, D. W., de Wet, J. R., Helinski, D. R., Howell, S. H., Wood, K. V. & de Luca, M. Transient and stable expression of the firefly luciferase gene in plant cells and transgenic plants. *Science*, 234(4778): 856-859. November 1986.

Ponce, E., Díaz E., & Padilla, F. Evolutionary Algorithm based on a Markov graphical model selection of promising solutions. En: *International Conference on Computing*. Mexico City, Mexico, 2004, pp. 3-9.

Schaffer, J. D. & Eshelman, L. J. On crossover as an evolutionary viable strategy. En: *Proceedings of the 4th International Conference on Genetic Algorithms*. Morgan Kaufmann, 1991, pp. 61-68.

Villalobos, V. *Los Transgénicos: oportunidades y amenazas*. Mundi-Prensa (Madrid, España). 2007.

[Título en Times New Roman y Negritas No.14]

Apellidos en Mayusculas -1er Nombre de Autor †, Apellidos en Mayusculas -2do Nombre de Autor

Correo institucional en Times New Roman No.10 y Cursiva

(Indicar Fecha de Envío: Mes, Día, Año); Aceptado(Indicar Fecha de Aceptación: Uso Exclusivo de ECORFAN)

Resumen

Título

Objetivos, metodología

Contribución

(150-200 palabras)

Abstract

Title

Objectives, methodology

Contribution

(150-200 words)

Keywords

Indicar (3-5) palabras clave en Times New Roman y Negritas No.11

Cita: Apellidos en Mayúsculas -1er Nombre de Autor †, Apellidos en Mayusculas -2do Nombre de Autor. Título del Paper. Título de la Revista. 2015, 1-1: 1-11 – [Todo en Times New Roman No.10]

*Correspondencia al Autor (Correo electrónico:)

† Investigador contribuyendo como primer autor.

Introducción

Texto redactado en Times New Roman No.12, espacio sencillo.

Explicación del tema en general y explicar porque es importante.

¿Cuál es su valor agregado respecto de las demás técnicas?

Enfocar claramente cada una de sus características

Explicar con claridad el problema a solucionar y la hipótesis central.

Explicación de las secciones del artículo

Desarrollo de Secciones y Apartados del Artículo con numeración subsecuente

[Titulo en Times New Roman No.12, espacio sencillo y Negrita]

Desarrollo de Articulos en Times New Roman No.12, espacio sencillo.

Inclusión de Graficos, Figuras y Tablas-Editables

En el *contenido del artículo* todo gráfico, tabla y figura debe ser editable en formatos que permitan modificar tamaño, tipo y número de letra, a efectos de edición, estas deberán estar en alta calidad, no pixeladas y deben ser notables aun reduciendo la imagen a escala.

[Indicando el titulo en la parte inferior con Times New Roman No.10 y Negrita]

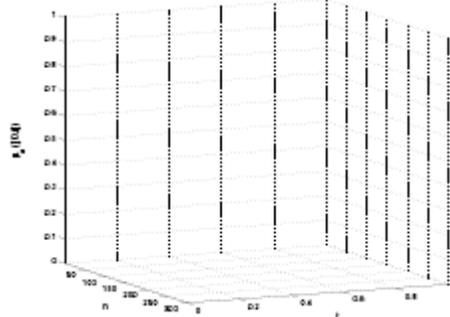


Grafico 1 Titulo y Fuente (en cursiva).

No deberan ser imágenes- todo debe ser editable.

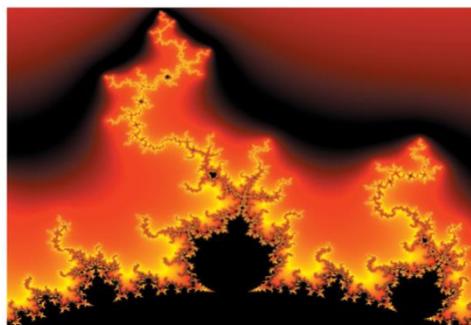


Figura 1 Titulo y Fuente (en cursiva).

No deberan ser imágenes- todo debe ser editable.

Tabla 1 Titulo y Fuente (en cursiva).

No deberan ser imágenes- todo debe ser editable.

Cada artículo deberá presentar de manera separada en **3 Carpetas**: a) Figuras, b) Gráficos y c) Tablas en formato .JPG, indicando el número en Negrita y el Titulo secuencial.

Para el uso de Ecuaciones, señalar de la siguiente forma:

$$Y_{ij} = \alpha + \sum_{h=1}^r \beta_h X_{hij} + u_j + e_{ij} \quad (1)$$

Deberán ser editables y con numeración alineada en el extremo derecho.

Metodología a desarrollar

Dar el significado de las variables en redacción lineal y es importante la comparación de los criterios usados

Resultados

Los resultados deberán ser por sección del artículo.

Anexos

Tablas y fuentes adecuadas.

Agradecimiento

Indicar si fueron financiados por alguna Institución, Universidad o Empresa.

Conclusiones

Explicar con claridad los resultados obtenidos y las posibilidades de mejora.

Referencias

Utilizar sistema APA. **No** deben estar numerados, tampoco con viñetas, sin embargo en caso necesario de numerar será porque se hace referencia o mención en alguna parte del artículo.

Ficha Técnica

Cada artículo deberá presentar un documento Word (.docx):

Nombre de la Revista

Título del Artículo

Abstract

Keywords

Secciones del Artículo, por ejemplo:

1. *Introducción*
2. *Descripción del método*
3. *Análisis a partir de la regresión por curva de demanda*
4. *Resultados*
5. *Agradecimiento*
6. *Conclusiones*
7. *Referencias*

Nombre de Autor (es)

Correo Electrónico de Correspondencia al Autor

Referencias

Formato de Originalidad



Sucre, Chuquisaca a ____ de ____ del 20 ____

Entiendo y acepto que los resultados de la dictaminación son inapelables por lo que deberán firmar los autores antes de iniciar el proceso de revisión por pares con la reivindicación de ORIGINALIDAD de la siguiente Obra.

Artículo (Article):

Firma (Signature):

Nombre (Name)

Formato de Autorización



Sucre, Chuquisaca a ____ de ____ del 20 ____

Entiendo y acepto que los resultados de la dictaminación son inapelables. En caso de ser aceptado para su publicación, autorizo a ECORFAN-Bolivia a difundir mi trabajo en las redes electrónicas, reimpresiones, colecciones de artículos, antologías y cualquier otro medio utilizado por él para alcanzar un mayor auditorio.

I understand and accept that the results of evaluation are inappealable. If my article is accepted for publication, I authorize ECORFAN-Bolivia to reproduce it in electronic data bases, reprints, anthologies or any other media in order to reach a wider audience.

Artículo (Article):

Firma (Signature)

Nombre (Name)

Revista de Análisis Cuantitativo y Estadístico

Calibración y selección del modelo de aprendizaje no supervisado K-Medias, de una encuesta sobre factores de riesgo en el consumo de drogas entre estudiantes

MEDINA-VELOZ, Gricelda, LUNA-ROSAS, Francisco Javier, TAVAREZ-AVENDAÑO, Juan Felipe y NARVAEZ-MURILLO, René Ulises

Diseño de algoritmos en labview en un posicionador xyz para autoequipamiento y apoyo didáctico

AGUILERA-HERNÁNDEZ, Martha Isabel, ORTIZ-SIMÓN, José Luis, RODRIGUEZ-MARTÍNEZ, Daniel y ROJO-VELASQUEZ, Gustavo

Efecto del raleo de frutos en el rendimiento de jitomate (*Lycopersicon esculentum* Mill). Bajo invernadero

GAYTÁN-RUELAS, Marina, VARGAS-ESPINOZA, Everardo, RIVERA-ARREDONDO, Marisa y MORALES-FÉLIX, Verónica de Jesús

Estudio de métodos difusos para la agrupación en un conjunto de datos

FUENTES, Juan Jaime

Universidad Politécnica del Valle del Évora

Evaluación del desempeño docente en la LMS Chamilo, de las ingenierías ejecutivas de la Universidad Tecnológica Fidel Velázquez: acciones de profesionalización

HERNÁNDEZ-CRUZ, María Guadalupe, ROMERO-ROJAS, Ruth Marcela, ORTIZ-ARANGO, Víctor Tonatiuh y VARGAS-GUTIÉRREZ, Luis Daniel

Universidad Tecnológica Fidel Velázquez

Sistema de programación dinámica de trayectorias para asistencia de nodos bajo demanda en redes inalámbricas de sensores

GONZÁLEZ-SILVA, Marco Antonio, FRANCO-MORENO, Juan José, BAEZA-REGALADO, Michelle y MORENO-ZAMORA, Brenda Paola

Universidad Politécnica Metropolitana de Hidalgo

Un algoritmo transgénico vs algoritmos de optimización

CORREA-VILLALÓN, Christian José, LUNA-RAMÍREZ, Enrique, SORIA-CRUZ, Jorge y CRUZ-VALENZUELA, Roberto

Instituto de Educación de Aguascalientes

Instituto Tecnológico El Llano Aguascalientes

