



ANÁLISIS DE DATOS DE LOS ENTORNOS VIRTUALES DE APRENDIZAJE, USANDO TÉCNICAS DE MINERÍA DE DATOS.

Bernabé Ortega, Rodrigo Aguilar, Viviana Quevedo

Facultad de Sistemas Mercantiles, Universidad Autónoma de los Andes Puyo

Facultad de Sistemas Mercantiles, Universidad Autónoma de los Andes Puyo

Facultad de Ciencias de la tierra, Universidad Estatal Amazónica

bernabe.ortega@gmail.com

Aguilar_rod@yahoo.com

vquevedo@uea.du.ec

Puyo, 30 de julio de 2015

ANÁLISIS DE DATOS DE LOS ENTORNOS VIRTUALES DE APRENDIZAJE, USANDO TÉCNICAS DE MINERÍA DE DATOS.

Resumen

Con el propósito de encontrar información relativa a la experiencia del uso de la tecnología como soporte al proceso de enseñanza aprendizaje mediante los Entornos Virtuales de Aprendizaje EVA, es necesario establecer un mecanismo de estudio de los datos que permita realizar un análisis de los mismo. La minería de datos a través de WEKA permite conocer con tendencias de comportamiento en los datos de entrada debido a su amplia gama de algoritmos y su potencialidad a la hora de pre-procesar y acreditar la información de origen, basando su estudio en descriptores reales.

Palabras Clave: Descriptores, Entorno Virtual de Aprendizaje, Weka

Abstract

In order to find information on the experience above using technology to support teaching and learning process through the Virtual Learning Environments EVA, it is necessary to establish a mechanism study of data that allows for an analysis of the same. Data mining through WEKA provides insight into behavioral trends in the input data due to its wide range of algorithms and its potential when it comes to pre - process and prove the source information, basing its study on real descriptors

Keywords: Descriptors, Virtual Environment Learning, Weka

Introducción

En el Ecuador, debido a los constantes cambios anunciados por el Consejo de Educación Superior CES, sobre los procesos de acreditación universitaria y evaluación de las carreras, se avizora, la opción de estudios semipresenciales como un medio de

educación alternativa en ciencias básicas, y humanísticas, por lo tanto, el uso de la tecnología educativa se convierte en un gran aliado para conseguir este objetivo. En este caso en la Universidad Regional Autónoma de los Andes UNIANDÉS, extensión Puyo, utiliza como instrumento de apoyo en el proceso de aprendizaje los Entornos Virtuales de Aprendizaje EVAs implementados sobre la plataforma Moodle, tras lo cual existe el acompañamiento docente.

La base de datos institucional de los EVAs contiene mucha información sobre la experiencia de uso de la tecnología como medio de apoyo en el proceso docente, por lo cual es pertinente hacer un estudio sobre los datos cuyos resultados determinen un plan de acciones a seguir que tributen en los objetivos y planes de acreditación de las carreras. En consecuencia, el análisis de datos se lo debe determinar mediante diferentes herramientas que permitan obtener resultados a partir de un grupo grande de datos altamente filtrados, tratados, y procesados que brinden un alto grado de confianza. La minería de datos se convierte en un aliado estratégico por su amplia gama de algoritmos de procesamiento de información, clasificación, clusterización, asociación, selección de atributos y características adicionales como pre-procesamiento de información, y revisión gráfica de los resultados, entre otros.

Las asignaturas se encuentran clasificadas como básicas, humanísticas y de especialidad, cada una tiene una significativa importancia en el proceso de formación de los profesionales, cada asignatura pertenece a un solo grupo y posee un EVA asignado dentro de la plataforma Moodle, que por defecto contiene el listado de estudiantes matriculados en la determinada asignatura. Desde aquí se gestiona el proceso de aprendizaje autónomo de los participantes.

Materiales Y Métodos

La recolección de los datos que intervienen en el proceso de análisis, se realizó mediante la aplicación de sentencias SQL sobre el motor de base de datos MySQL institucional que contiene los registros históricos de la actividad de la plataforma

Moodle como una herramienta complementaria a la labor docente y de aprendizaje autónomo de los estudiantes. El resultado de las consultas es un conjunto aproximado de 720000 registros filtrados que se han procesado como archivos separados por comas CSV con la ayuda de software gratuito para la estandarización y depuración de la información tales como: PhpMyAdmin como herramienta gratuita escrita en php con la finalidad de administrar MySql desde el web; CSVed 2.3.2 (freeware Windows csv file editor) de Sam Francke 2004-2014 que permite evaluar la validez de la estructura de datos (número de columnas) y reparación de los mismos; EditPad Lite 7 de Jan Goyvaerts 1996-2015 cuya funcionalidad permite manejar archivos con grandes volúmenes de datos, Weka Gui Chooser de Waikato Enviroment for Knowledge Analysis 199-2014 como software que contiene una colección de herramientas y algoritmos para análisis y modelos predictivos.

| Query | Registros | T. ejecución (sec) | Agrupación |
|--|------------------|---------------------------|-------------------|
| Actividad en los EVA por los estudiantes | 549 | 5.638 | Nombre de curso |
| Actividad en los EVA por los docentes | 948 | 0.745 | Id de curso |
| Actividades por intervalos de fechas | 3013 | 12.31 | Fecha |
| Uso de foros dentro del aula | 42 | 0.616 | Id curso, docente |
| Uso de recursos en el aula | 82 | 0.027 | Curso |
| Estudiantes activos por módulo de curso | 85 | 1.600 | Id curso, docente |
| Visitas a los cursos | 177 | 0.307 | Id curso |
| Actividad docente en línea | 68 | 0.352 | Id curso |
| Puntuación estudiante tema modulo | 13234 | 0.550 | Estudiante |
| Login por actividad | 370 | 14.195 | Estudiante |

| | | | |
|---|-----|-------|----------|
| Participación semanal del instructor | 150 | 0.647 | Id curso |
| Uso de recursos por categoría | 60 | 0.136 | Curso |
| Variedad de recursos docente dentro del EVA | 29 | 5.965 | Docente |
| Reporte general de composición de un EVA | 179 | 0.414 | Fechas |
| Ultimo acceso | 83 | 37971 | Docente |

Tabla 1: Filtro y selección de la información

Los resultados que muestra la tabla 1, son analizados en cuanto a su estructura, contenido y consistencia, luego de lo cual ingresa a un proceso de transformación de la data generando conjuntos de datos procesados localmente y categorizados de acuerdo a sus atributos e instancias; este proceso obtiene los siguientes resultados.

| Dataset | Atributos | Instancias |
|------------------------|------------------|-------------------|
| dtstActividadCurso | 3 | 549 |
| dtstActividadPorCurso | 4 | 549 |
| dtstUsoEstudiantes | 4 | 42 |
| dtstCursoActivo | 5 | 85 |
| dtstInformeEstudiantes | 5 | 13234 |
| dtstRecursosCategoria | 5 | 60 |
| dtstReporteCurso | 6 | 179 |

Tabla 2: Atributos e instancias de los conjuntos de datos

Los datasets de la tabla 2, son tratados durante la etapa de pre-proceso mediante filtros supervisados aplicados como discretos y selección de atributos, para luego pasar por la etapa de aprendizaje e interpretación y evaluación.

El dataset dtstUsoEstudiantes de extensión arff, contiene una estructura formada por la definición de una relación denominada usoestudiantes la información de 4 atributos

numéricos, 42 instancias y una clase que contiene 35 descriptores correspondientes a las asignaturas más representativas de la muestra, donde se aplica un algoritmo supervisado de aprendizaje, orientando el experimento basado en la usabilidad y actividad de los módulos de un EVA, por parte de los estudiantes dentro de una plataforma virtual.

Resultados

Bajo el enfoque de determinar la eficacia de los temas propuestos en los EVAs para los estudiantes, es necesario evaluar los datos en vectores característicos denominados descriptores, y dado que ya se posee la estructura necesaria para su procesamiento, sobre los datos de entrada, se aplica el Clasificador KNN, que en Weka se denomina IBk como miembro del conjunto lazy, bajo los argumentos del algoritmo por defecto. La data es procesada tanto para el nivel de entrenamiento como para las pruebas.

```
Classifier output
=== Run information ===

Scheme:weka.classifiers.lazy.IBk -K 1 -W 0 -A "weka.core.neighboursearch.Linear
Relation: CantidadForos-weka.filters.supervised.attribute.Discretize-Rfirst
Instances: 545
Attributes: 5
          modulos
          estudiantes
          actividad
          accion
          class
Test mode:evaluate on training data

=== Classifier model (full training set) ===

IB1 instance-based classifier
using 1 nearest neighbour(s) for classification

Time taken to build model: 0 seconds

=== Evaluation on training set ===
=== Summary ===

Correctly Classified Instances      506           92.844 %
Incorrectly Classified Instances    39            7.156 %
Kappa statistic                    0.9247
Mean absolute error                 0.0051
Root mean squared error             0.0479
Relative absolute error             9.407 %
Root relative squared error         29.0965 %
Total Number of Instances          545
```

Gráfico 2. Valores obtenidos en modo entrenamiento KNN


```
Classifier output

=== Run information ===

Scheme:weka.classifiers.lazy.IBk -K 1 -W 0 -A "weka.core.neighboursearch.LinearNNSearch -A \
Relation: CantidadForos-weka.filters.supervised.attribute.Discretize-Rfirst-last
Instances: 545
Attributes: 5
          modulos
          estudiantes
          actividad
          accion
          class
Test mode:10-fold cross-validation

=== Classifier model (full training set) ===

IB1 instance-based classifier
using 1 nearest neighbour(s) for classification

Time taken to build model: 0 seconds

=== Stratified cross-validation ===
=== Summary ===

Correctly Classified Instances      501           91.9266 %
Incorrectly Classified Instances    44            8.0734 %
Kappa statistic                    0.915
Mean absolute error                 0.0057
Root mean squared error             0.051
Relative absolute error             10.4677 %
Root relative squared error         30.9364 %
Total Number of Instances          545
```

Gráfico 4: KNN mediante Cross-validation con 10 muestras

El gráfico 4 fue obtenido a partir del dataset que es motivo de la investigación evaluado con 10 muestras randomicas, y obteniendo un alto grado de confianza determinado por el 91.9266% de instancias correctamente clasificadas.

| | | | | |
|---------|---------------------------|---------|--------|----|
| K=4 w=0 | Cross-validation 10 folds | 84.7706 | 24 | 58 |
| K=4 w=0 | Percentage Split 75% | 66.9118 | 24, 27 | 10 |

Tabla 2: Valores de pruebas con aumentos de K en opciones de test

La tabla 2 muestra los resultados obtenidos para diferentes valores de k en correspondencia con su grado de efectividad

Discusión

La aplicación de distintos valores de k sobre las opciones de: training, cross validation, percentage Split al 75% (75% de entrenamiento, 25% para pruebas) muestran la clara tendencia de resaltar al descriptor 24 que corresponde en el dataset a una asignatura de especialización dentro de la carrera, pese a que la aplicación de la opción de test cross validation presente diferentes valores debido a como se trata la muestra, siempre la tendencia apunta al descriptor antes mencionado, con un porcentaje de acierto altamente confiable con una media de 86.21%.

Los valores de w corresponden al peso que se puede asignar a los vecinos, en tanto k es el número cercano de vecinos con los que se va a clasificar, sin embargo no se realiza ningún experimento con los valores de w debido a que las asignaturas que están representadas en el dataset corresponden a la clasificación básica, humanística y de especialización, por lo tanto no se comparan entre ellas.

Los datos correspondientes a los porcentajes de éxito se encuentran expresados en la tabla 2 y se identifican como peso, y son obtenidos de la matriz de confusión que arroja cada uno de los muestreos, sin embargo a mayor valor de k, la tendencia ya no es única sino que muestra un nuevo valor correspondiente a otra asignatura clasificada en el grupo de especialización.

Conclusiones

Los datos obtenidos expresan, que la representación de las aulas virtuales tiene mayor efectividad entre los estudiantes de las asignaturas cuyas categorías corresponden a los temas de especialización en la carrera. Los resultados obtenidos son altamente confiables, así lo demuestra la media calculada sobre el porcentaje de efectividad de las diferentes muestras que la marcan como tendencia sobre el 86.21% sobre el mismo descriptor.

Para el cumplimiento del proceso estricto del análisis de datos, los datos deben ser seleccionados, obtenidos, pre-procesados, transformados y enviados como data de entrada a la máquina de aprendizaje, aquí es donde se aplican los diversos algoritmos y realizar experimentos de acuerdo a la parametrización de sus argumentos, en este caso se define una tendencia específica sobre uno de los descriptores del grupo de resultados, que a su vez corresponde a una asignatura del grupo de especialidad.

Sin importar los valores de K y los distintos experimentos, se puede concluir que la tendencia es la misma gracias al alto nivel de efectividad del resultado del procesamiento de la data o en su defecto de los algoritmos.

Bibliografía

- a2.-Estatuto-Orgánico-por-procesos-Febrero-2014.pdf. (s. f.). Recuperado a partir de <http://www.ceaaces.gob.ec/sitio/wp-content/uploads/2014/02/a2.-Estatuto-Orga%CC%81nico-por-procesos-Febrero-2014.pdf>
- Classification via Decision Trees in WEKA. (s. f.). Recuperado 15 de junio de 2015, a partir de <http://facweb.cs.depaul.edu/mobasher/classes/ect584/WEKA/classify.html>
- CSV file editor, for Windows. (s. f.). Recuperado 15 de junio de 2015, a partir de <http://csved.sjfrancke.nl/>

- Diapositiva 1 - Resultados_principales_140515.Tic.pdf. (s. f.). Recuperado a partir de http://www.ecuadorencifras.gob.ec/documentos/web-inec/Estadisticas_Sociales/TIC/Resultados_principales_140515.Tic.pdf
- EditPad Lite - Free Text Editor for Windows. (s. f.). Recuperado 15 de junio de 2015, a partir de <http://www.editpadlite.com/>
- Ejemplo de Aplicación de Minería de datos a WEKA - YouTube. (s. f.). Recuperado 15 de junio de 2015, a partir de <https://www.youtube.com/watch?v=Wz5OsY72CQA>
- El uso de Internet en Ecuador creció 11 veces en siete años | El Comercio. (s. f.). Recuperado 15 de junio de 2015, a partir de <http://www.elcomercio.com/tendencias/ecuador-internet-datos-tecnologia-usuarios.html>
- Indicadores de desempeño, informe de gestión, cumplimiento de metas - Indicadores de desempeño, informe de gestión, cumplimiento de metas | CES - Consejo de Educación Superior | Ecuador. (s. f.). Recuperado 15 de junio de 2015, a partir de <http://www.ces.gob.ec/transparencia/informacion-de-planificacion-institucional/indicadores-de-desempeno-informe-de-gestion-cumplimiento-de-metas>
- LOS ENTORNOS VIRTUALES DE APRENDIZAJE | Suárez Guerrero | Investigación Educativa. (s. f.). Recuperado 15 de junio de 2015, a partir de <http://revistasinvestigacion.unmsm.edu.pe/index.php/educa/article/view/3776>
- Machine Learning Project at the University of Waikato in New Zealand. (s. f.). Recuperado 15 de junio de 2015, a partir de <http://www.cs.waikato.ac.nz/ml/index.html>
- Modelo genérico para la evaluación de las carreras presenciales y semipresenciales - CEAACES. (s. f.). Recuperado 15 de junio de 2015, a partir de <http://www.ceaaces.gob.ec/sitio/modelo-generico-de-carreras-presenciales-y-semipresenciales/>
- phpMyAdmin. (s. f.). Recuperado 15 de junio de 2015, a partir de http://www.phpmyadmin.net/home_page/index.php

- Software Libre | Secretaría Nacional de la Administración Pública. (s. f.). Recuperado 15 de junio de 2015, a partir de <http://www.administracionpublica.gob.ec/software-libre/>
- weka j48 classification tutorial - YouTube. (s. f.). Recuperado 15 de junio de 2015, a partir de <https://www.youtube.com/watch?v=tIBf2FpXGeY>