

Capítulo 10

Diseño e Implementación de un Módulo de Analítica de Aprendizaje, y su Aplicación para la Evaluación de Experiencias Educativas

José A. Ruipérez-Valiente, Pedro J. Muñoz-Merino, *Senior Member, IEEE* y Carlos Delgado Kloos, *Senior Member, IEEE*

Title — Design and Implementation of a Learning Analytics Module, and its Application for Evaluating Educational Experiences

Abstract — Most e-learning platforms are able to collect large datasets of students' interactions as events; however that data is difficult to be interpreted directly by learning stakeholders. In this work we unify and connect several of our previous research studies giving a general context of our learning analytics research on Khan Academy. We propose a set of interesting indicators in order to learn more about the learning process. Furthermore, we have designed and implemented a learning analytics module called ALAS-KA which displays individual and class visualizations for these parameters. Finally we make use of ALAS-KA and the parameters to evaluate learning experiences.

Index Terms — Data visualization, distance learning, learning analytics, systems engineering education

Abstract — La mayoría de plataformas de aprendizaje virtual recopilan grandes cantidades de datos, pero éstos son difícilmente interpretables directamente por los actores que intervienen en el aprendizaje. En este trabajo unificamos y conectamos diversos trabajos que hemos realizado, dando un contexto general de nuestra investigación en analítica de aprendizaje sobre la plataforma Khan Academy. Para ello, se proponen una serie de indicadores relacionados con el proceso de aprendizaje de los estudiantes y cómo implementarlos. Además, se ha desarrollado la aplicación de analítica de aprendizaje ALAS-KA. Finalmente se dan ejemplos de cómo usar dichos indicadores y ALAS-KA para evaluar experiencias educativas.

Index Terms— Analítica de aprendizaje, aprendizaje a distancia, sistemas para educación a la ingeniería, visualización de datos

Diferentes partes de este trabajo han sido parcialmente presentadas originalmente en otras publicaciones [13]–[17].

José A. Ruipérez-Valiente, Pedro J. Muñoz-Merino and Carlos Delgado Kloos pertenecen al Departamento de Ingeniería Telemática, Universidad Carlos III de Madrid, Avenida Universidad 30, 28911 Leganés (Madrid) España. E-mail: {jruipe, pedmume, cdk}@it.uc3m.es

José A. Ruipérez-Valiente también pertenece a IMDEA Networks Institute, Av. del Mar Mediterráneo 22, 28918 Leganés (Madrid) España.

I. INTRODUCCIÓN

EN los últimos años ha habido un incremento de la Educación a distancia utilizando plataformas de aprendizaje virtual, ya sea mediante cursos completamente online o combinando educación presencial y online. Las plataformas denominadas LMSs (Learning Management Systems) como Moodle, WebCT, dotLRN o Dokeos han sido las mayormente empleadas para contener los cursos y clases virtuales.

En la actualidad, ha surgido con fuerza el término MOOC (Massive Online Open Course). Estos cursos masivos abiertos online, se diferencian de lo que ya venía siendo habitual desde hace tiempo por distintas universidades, precisamente en ese componente que aglutina a tanta gente al mismo tiempo: un mismo curso al que cualquiera se puede suscribir, bajo un calendario, siendo tomado por miles de alumnos simultáneamente en todas las partes del mundo. Algunas de las plataformas que han surgido para dar respuesta a este nuevo paradigma son Coursera, Udacity, edX, MiriadaX o Khan Academy.

Entre las necesidades que surgen con la aparición de los MOOCs, se acrecienta la de tener una analítica de aprendizaje (learning analytics) potente, que permita realizar un seguimiento automático detallado de los alumnos y así poder supervisarlos de forma sencilla. Este aumento de necesidad de herramientas automáticas para evaluar el proceso de aprendizaje se justifica debido a la gran cantidad de alumnos que hay en estos cursos.

Nuestra motivación en este trabajo es conectar distintas investigaciones que hemos realizado con anterioridad con el fin de dar un flujo completo al proceso de analítica de aprendizaje que hemos seguido en Khan Academy. Para ello se revisaran una serie de parámetros de alto nivel que transforman los datos generados por los estudiantes para poder mejorar el proceso de aprendizaje, lo cual es un gran reto. La clave se encuentra en determinar la información que resulta de utilidad para mejorar el proceso de aprendizaje, y cómo es posible conseguirla mediante una transformación de los datos que se encuentran disponibles. Dicha información y la forma de obtenerla depende en muchos casos del tipo de plataforma a considerar: su semántica,

funcionalidades concretas o el tipo de uso que se le vaya a dar al curso.

En este trabajo nos gustaría ser capaces de responder a preguntas como por ejemplo: ¿Qué usuario ha sido más eficiente?, ¿Está motivado por los elementos de ludificación?, ¿En qué horario del día trabaja mejor?, ¿Ha abusado de pistas?, ¿Qué alumno ha sido más constante? Este conocimiento puede proporcionar las vías a los profesores para ser capaces de actuar antes de que un estudiante abandone el curso, de realizar recomendaciones personalizadas en función de los resultados de cada estudiante, o adaptar los contenidos del curso en función de las preferencias y hábitos de los usuarios. Así mismo, puede proporcionar a los alumnos un medio de autorreflexión de su aprendizaje para intentar cambiar determinados hábitos y para mejorar su motivación hacia el aprendizaje virtual.

Por ello en este trabajo se plantean 3 objetivos principales, que se encuentran conectados para completar el proceso de analítica de aprendizaje que hemos seguido:

1. En primer lugar definir una serie de parámetros que transforman los datos de bajo nivel que contiene la base de datos de Khan Academy en información de alto nivel que ayuda a conocer más acerca del proceso de aprendizaje y se ha establecido la manera de procesarlos con fórmulas concretas.
2. En segundo lugar el desarrollo de ALAS-KA, que es la aplicación que implementará dichos parámetros de alto nivel que serán obtenidos a través de algoritmos que procesaran los datos generados en Khan Academy. ALAS-KA es una herramienta de analítica de aprendizaje que implementará un panel de visualizaciones que podrá ser accedido por los profesores y alumnos cuando estén interactuando con un curso de Khan Academy. Los parámetros procesados serán transmitidos a estudiantes y alumnos mediante visualizaciones individuales y de clase.
3. En tercer lugar se utilizará este módulo y los parámetros definidos para evaluar el proceso de aprendizaje, ilustrándolo en los cursos 0 de la Universidad Carlos III de Madrid que son atendidos por los estudiantes de primer año y en los cuáles se usa Khan Academy como soporte. El módulo ALAS-KA y sus parámetros se utilizará para la evaluación de dichas experiencias con estudiantes reales y las conclusiones obtenidas servirán para ver cuáles han sido los resultados, mejorar el proceso de aprendizaje de estudiantes y profesores así como poder detectar problemas y solucionarlos de cara a los cursos en los siguientes años.

II. ESTADO DEL ARTE

En la actualidad la mayoría de los servicios y herramientas han sido informatizados y se están generando más datos que nunca. Algunos informes como el de IBM [1] estiman que el 90% de los datos existentes han sido creados en los últimos dos años debido al *boom* que ha habido en el sector.

En el sector del aprendizaje virtual también se hace uso de los datos generados para extraer información que sea de utilidad para profesores, alumnos y otros agentes involucrados en el aprendizaje. De acuerdo a la 1ª Conferencia Internacional en Analítica de Aprendizaje y

Conocimiento [2], analítica de aprendizaje se describe como “*la medida, recolección, análisis e informe de los datos sobre los estudiantes y sus contextos, con el propósito de entender y optimizar el aprendizaje y el entorno en el que ocurre*”. En el ámbito del aprendizaje virtual hay varios términos relacionados con “learning analytics” como las “academics analytics” o el “educational data mining”. El término “academics analytics” está más relacionado con la explotación de grandes bases de datos educacionales a nivel institucional [3], mientras que “learning analytics” se centra más en el proceso de aprendizaje del alumno.

El ciclo que engloba el proceso de analítica de aprendizaje es un tema que ha sido abordado en diferentes publicaciones. Por ejemplo, se puede definir como un proceso con 5 pasos que son Captura, Informar, Predecir, Actuar y Refinar [4]. Otro trabajo lo define como un ciclo en el que los estudiantes generan datos y con esos datos se procesan unas métricas, dónde el paso clave es cerrar el ciclo con intervenciones y *feedback* sobre esos resultados para mejorar el proceso [5]. Esta última aproximación se asemeja más a la forma en la que nosotros aplicamos analítica de aprendizaje.

Otros trabajos también han abordado el procesado de datos para obtener información inteligente, por ejemplo para detectar los comportamientos que hace el usuario al intentar resolver un problema [6] o para intentar estimar el rendimiento y puntuación obtenido en un examen antes de que el alumno lo realice [7].

También se pueden encontrar en la literatura herramientas de visualización similares a ALAS-KA. Por ejemplo para soportar la monitorización del trabajo en un entorno personal de aprendizaje con la herramienta CAMera [8]. Otras dos herramientas muy similares a ALAS-KA son en primer lugar CourseVis [9], que es una plataforma de visualizaciones que realiza un seguimiento de los datos creados en WebCT y Moodog [10] que está implementada como un plug-in para explotar los datos generados en Moodle.

Se pueden encontrar otras posibles herramientas basadas en analítica de aprendizaje, por ejemplo sistemas recomendadores para aconsejar recursos a los estudiantes [11] o para la formación de grupos de trabajo en base a estudiantes que tengan características similares [12].

III. PROPUESTA DE PARÁMETROS DE ALTO NIVEL

Esta sección se centra en el diseño propuesto de los parámetros de alto nivel a partir de los datos de bajo nivel generados por el entorno educativo. El proceso para el diseño de estas medidas conlleva tanto un análisis exhaustivo de los datos disponibles en la base de datos, así como establecer cuáles son las informaciones relacionadas con el aprendizaje que se pueden y se desean alcanzar. Los distintos parámetros propuestos están agrupados por funcionalidad en 5 bloques. Información más detallada sobre el cálculo de dichos parámetros ha sido publicada con anterioridad [13], [14].

A. Progreso Total en la Plataforma

Los parámetros pertenecientes a este bloque se van a centrar en describir cuánto ha interactuado el usuario con la plataforma. No van a valorar si el usuario lo ha hecho bien o mal, sino la cantidad de interacción total, ejercicios y videos

accedidos, tiempo empleado, etcétera. El objetivo de estas medidas es el de dar una idea de si el usuario ha participado mucho o poco con la plataforma al margen de que lo haya hecho bien o mal

- **Uso efectivo de la plataforma:** El objetivo de esta medida es dar una idea global de cuanto ha usado un usuario la plataforma en relación a los recursos disponibles que se encuentran en ella (en este caso vídeos y ejercicios).
- **Abandono de vídeos o ejercicios:** Esta medida está relacionada con si el usuario deja sin terminar los recursos que comienza. Para el caso de los vídeos sería la búsqueda de usuarios que comienzan vídeos pero no lo llega a terminar. Para el caso de los ejercicios se propone que sea aquellos usuarios que comienzan un ejercicio pero no llegan a alcanzar la *proficiency*.
- **Tiempo total en la plataforma:** Cálculo del tiempo empleado por el estudiante en ejercicios y vídeos en la durante su interacción con la plataforma.
- **Uso de elementos opcionales:** La plataforma Khan Academy también incluye algunos elementos opcionales que no son de uso obligatorio para realizar ejercicios o ver videos ni son necesarios para superar el curso. En esta medida se analiza qué usuarios hacen uso de ellos con una ponderación [15].
- **Focalizador de vídeos o ejercicios:** Se analiza el tipo de recursos en el cual focaliza el usuario más su interacción en proporción, puede indicar qué estrategia de aprendizaje sigue cada usuario.

B. Progreso Correcto en la Plataforma

El objetivo de este bloque de medidas es analizar el progreso correcto del usuario dentro de la plataforma. Al contrario del punto anterior, dónde sólo medíamos la cantidad de uso, en este bloque sí se valora lo bien que lo haya hecho el usuario. Estas medidas se dividen principalmente en dos tipos: relacionadas con el progreso que ha hecho el usuario tanto en ejercicios como en vídeos, y relacionadas con la eficiencia del usuario progresando en los vídeos y ejercicios:

- **Efectividad en ejercicios:** Función específica adaptada al cálculo del progreso efectivo en ejercicios que tiene en consideración el entorno y especificidades de Khan Academy [16].
- **Efectividad en vídeos:** Función específica adaptada al cálculo del progreso efectivo en vídeos [16].
- **Eficiencia en ejercicios:** Se calcula una eficiencia para resolver ejercicios correctamente por primera vez, en esta medida se combina el número de intentos y el tiempo necesario para resolver correctamente el primer ejercicio de cada tipo.
- **Eficiencia en vídeos:** El concepto de eficiencia relacionada con videos sería relativo a aquellos videos que se han completado o se ha alcanzado un mínimo de porcentaje completado. En dichos vídeos, si el usuario ha repetido trozos de la visualización del video, es decir, si ha tenido que visualizarlo más de una vez habría que considerar que ha sido menos eficiente.

C. Distribución del Tiempo de Uso de la Plataforma

En esta sección se analizan los hábitos temporales de los estudiantes en su interacción con la plataforma. Una de las medidas más directas que podemos ver de este análisis es el tramo temporal (mañana, tarde o noche) en el que cada

usuario focaliza su aprendizaje. Esto también permite analizar con facilidad en que horario una clase tiene más usuarios trabajando. Además se pueden calcular eficiencias de cada usuario en cada uno de ellos, que en ocasiones permiten encontrar horarios en los que un estudiante puede trabajar mejor. También se puede analizar la constancia en el aprendizaje del usuario durante el progreso del curso.

- **Uso de la plataforma y eficiencia por intervalos horarios:** Esta medida se centra en dividir el día en tres intervalos que son de mañana, tarde y noche. Posteriormente se calcula el tiempo dedicado en ejercicios y vídeos en cada uno de estos intervalos para ver en cuál de ellos se focaliza el trabajo de un estudiante. También es posible utilizar dichos intervalos para calcular eficiencias resolviendo ejercicios bien en cada uno de los intervalos.
- **Media y varianza del tiempo de uso diario:** La media muestral representará una idea de la cantidad de trabajo que ha realizado el alumno a lo largo del curso. Por otra parte, cuanto mayor sea la varianza indicará que el usuario ha sido más inconstante con respecto de su media. Cuanto menor sea, implicará que su trabajo diario se encuentra más cercano a su valor medio, es decir habrá sido más constante.

D. Hábitos de Ludificación

Tal como se ha demostrado en múltiples estudios, los elementos de ludificación pueden ser buenos motivadores para los alumnos y ayudar en su aprendizaje. Conociendo que alumnos se motivan más ante este tipo de técnicas de ludificación, es posible los profesores puedan aplicarlas posteriormente si se considera adecuado. Por ello en esta sección los parámetros tratan de dar una medida de cuáles son los usuarios más motivados por los elementos de juegos:

- **Interés en ganar medallas:** Se puede establecer una simple comparación de la cantidad media de medallas que gana un estudiante cualquiera, con respecto de las que ha ganado el estudiante, se puede saber si ha conseguido más medallas que la media o no.
- **Proporción de puntos por ludificación:** Los *energy points* en la plataforma se pueden ganar mediante la resolución de ejercicios, visionado de vídeos o realizando elementos de ludificación como ganar medallas o completar *goals*. Establecemos una medida comparativa para ver qué porcentaje de esos puntos corresponde a elementos de ludificación, comparado con aquellos conseguidos mediante ejercicios y videos. Esta medida nos sirve para comparar unos estudiantes con otros y ver cuales ganan un mayor porcentaje de puntos debido a elementos de ludificación.

E. Hábitos Resolviendo Ejercicios

El flujo de eventos con las marcas temporales que realiza un usuario cuando está resolviendo ejercicios es almacenado en gran detalle por el sistema de Khan Academy. Esto permite posteriormente reconstruir este flujo y analizar cuáles son los hábitos de determinado estudiante al resolver ejercicios. De esta forma, se pueden detectar comportamientos que se consideran erróneos en base a determinadas casuísticas, así como otros perfiles que permitan conocer más en profundidad al alumno para posteriormente realizar recomendaciones o intervenciones:

- **Seguidor de recomendaciones:** El sistema monitoriza cuando el estudiante ha entrado a ejercicios a través de una recomendación del sistema. Esto permite conocer en qué porcentaje de las veces que un usuario accede a resolver un ejercicio, éste lo ha hecho proviniendo de una recomendación del sistema.
- **Evitador de vídeos:** En caso de que el usuario haya respondido incorrectamente y no haya visto el vídeo asociado al ejercicio.
- **Evitador de pistas:** En caso de que el usuario responda incorrectamente y no haya pedido pistas, si el ejercicio tenía pistas disponibles
- **Usuario no reflexivo:** En caso de que el usuario haya respondido incorrectamente antes de que hayan pasado 10 segundos o que los posteriores intentos de resolución hayan estado separados entre ellos por menos de 10 segundos también, se considerará que el usuario no ha reflexionado en sus respuestas.
- **Abuso de pistas:** En caso de que entre al ejercicio y pida una pista antes de 10 segundos o que la diferencia temporal entre pistas pedidas sea de menos de 10 segundos, se considerará que está abusando de pistas porque no ha tenido tiempo de reflexionar sobre el enunciado del problema o sobre la pista dada anteriormente.

IV. DESARROLLO DEL MÓDULO ALAS-KA DE ANALÍTICA DE APRENDIZAJE

A. Funcionalidad Requerida

La funcionalidad principal del sistema es poner a disposición de los usuarios visualizaciones que permitan recibir la información de alto nivel procesada por los parámetros propuestos. Es importante el uso de un diseño sencillo que permita la rápida interpretación de los resultados por cualquier usuario no técnico. El procesado de dichos datos tendrá que ser generado de forma periódica por la plataforma, para que las medidas se encuentren actualizadas cada vez que sean consultadas. Se plantea dos grupos de visualizaciones:

- Las primeras de forma individual centradas en cada usuario por separado y además con una comparación con el valor medio de la clase en dicha medida. Esto permite dar una idea de en qué punto se encuentra cada usuario y además compararlo con el resto de sus compañeros.
- Las segundas son visualizaciones globales de clase en forma de diagramas de tarta, esto permite ver en qué punto se encuentran la mayoría de los usuarios de una clase dando una visión global del conjunto.

Dentro del sistema va a haber dos tipos de usuarios, por una parte los profesores y por otra los alumnos. Por lo tanto también es necesario tener en cuenta esto a la hora de asignar privilegios para las visualizaciones. Los profesores tendrán acceso a las visualizaciones individuales de todos los usuarios así como a las visualizaciones de clase, mientras que los alumnos sólo podrán acceder a ver las visualizaciones individuales suyas para conocimiento propio y no podrán acceder a las de ningún otro alumno ni a las de clase. Además, el acceso a las visualizaciones estará restringido a sólo usuarios del sistema.

Otra de las funcionalidades que incluye el sistema, es una descripción verbal de cada usuario de forma individual para cada una de las medidas en forma de tabla. En dicha tabla se harán diferentes agrupamientos para cada una de las medidas y una descripción verbal informará de las características de cada usuario.

B. Arquitectura de ALAS-KA

En la figura 1 (adaptada de publicaciones previas [17], [18]) se muestra el diagrama del sistema integrado en el mismo servidor de Google App Engine con Khan Academy. Para una descripción más detallada de la arquitectura puede consultar referencias previas [17]. Ahora se van a comentar por separado cada uno de los elementos y cuál es su función, así como su interacción con otros elementos.

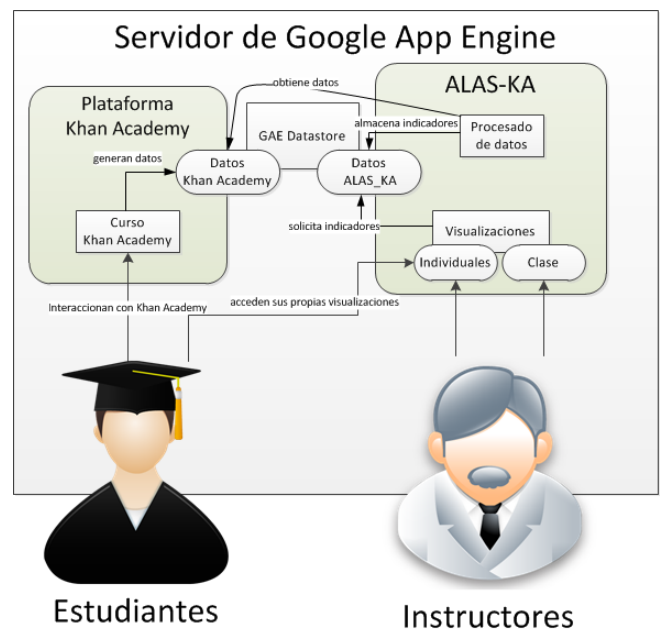


Figura 1. Arquitectura del Sistema ALAS-KA en el mismo servidor con Khan Academy.

1) Estudiantes

Los estudiantes son los que van a interactuar con la plataforma Khan Academy y generarán las entidades de bajo nivel de Khan Academy que luego son necesarias para reconstruir el flujo que han seguido los usuarios y aplicar todos los procesamientos para generar los parámetros de alto nivel. Además los estudiantes también pueden acceder a la aplicación ALAS-KA y recibir realimentación con las visualizaciones de ellos mismos para poder reflexionar sobre su propio aprendizaje, pero no podrán acceder a las visualizaciones de la clase o a las de otros usuarios.

2) Profesores

El trabajo de los profesores es el diseño de los vídeos y ejercicios de los cursos de Khan Academy, pero además la monitorización del trabajo de los alumnos durante el curso. ALAS-KA proporciona las herramientas necesarias para supervisar la progresión y el trabajo de cada alumno por separado mediante visualizaciones individuales así como la clase en conjunto mediante visualizaciones globales. Los profesores pueden acceder a todas ellas. Pueden intervenir en caso de que se observe alumnos con problemas, hacer advertencias de forma temprana, detectar ejercicios o vídeos

con problemas y en realizar recomendaciones a los estudiantes de forma más sencilla.

3) App Engine Datastore

El App Engine Datastore proporciona el almacenamiento para Khan Academy. La mayoría de los eventos que generan los estudiantes durante su proceso de aprendizaje son almacenados. Ya que ALAS-KA se ha diseñado en forma de plug-in para funcionar de forma conjunta en el mismo servidor de Khan Academy, la forma más sencilla es usar también este sistema para la persistencia de datos. Se puede encontrar por una parte las entidades de Khan Academy, que son por ejemplo las generadas por la interacción de los usuarios. Por otra parte las entidades de ALAS-KA que son usadas tanto para guardar parámetros de configuración como para almacenar los resultados de los indicadores procesados.

4) Visualizaciones

La API Google Charts es la elegida para generar las visualizaciones debido a su simplicidad pero a la vez su potente entorno de trabajo. Un punto importante es que las gráficas están generadas usando HTML5/SVG por lo que proporciona compatibilidad entre navegadores y entre dispositivos. Google Charts genera los gráficos en el lado del cliente por lo que el servidor no tiene que realizar ningún procesamiento adicional. En el caso de esta aplicación, los datos necesarios para las visualizaciones son pedidos al Datastore, y sólo son necesarios estos datos sin ningún procesamiento adicional para generar las visualizaciones.

C. Diseño del Módulo de Procesado de Datos

Uno de los retos en analítica de aprendizaje así como en otros dominios donde hay que hacer grandes extracciones y procesados de inmensas cantidades de datos, es la forma de hacerlo eficientemente y sin saturar el servidor.

Las consultas más pesadas son demasiado lentas como para ser ejecutadas en tiempo real ante la petición de un usuario. Este hecho hace que sea imposible realizar dicho procesamiento del indicador cada vez que una visualización sea solicitada. Para solucionar este problema se configura el sistema de App Engine Cron Service¹ para procesar los parámetros necesarios en intervalos de tiempo regulares. Una vez que el procesamiento ha terminado, se almacenan los resultados en la Datastore y pueden ser recuperados y visualizados rápidamente.

Se han creado cinco tareas *cron jobs* diferentes, uno para cada uno de los cinco grupos de medidas definidas en la sección III. El sistema lanzará cada cierto tiempo una petición HTTP GET a dicha URL para ejecutar el *cron job*. Dicho código generará una entidad *task* por cada alumno de la plataforma y éstas serán encoladas en su *queue* correspondiente. Una *task* está considerada como una unidad pequeña y discreta de trabajo en App Engine.

La Task Queue Python API² de App Engine, proporciona un poderoso sistema para ejecutar tareas en segundo plano. Se encarga de ejecutar las peticiones que no provienen del usuario. También se han generado cinco *queues* diferentes, una para cada uno de los *cron jobs*. Las *tasks* generadas por cada uno de los *cron jobs* serán encoladas en su *queue* correspondiente. Una vez añadidas las irá ejecutando en el mismo orden de llegada y cuando

sea posible en pequeñas unidades de trabajo que no saturarán el servidor.

V. EVALUACIÓN DEL PROCESO DE APRENDIZAJE EN EXPERIENCIAS EDUCATIVAS

La UC3M empezó a utilizar Khan Academy para preparar los cursos 0, que son los cursos pre-universitarios en los que se repasan los conocimientos que los alumnos deberían poseer antes de comenzar los cursos oficiales de Grado en ciencias tecnológicas. La idea es aplicar la metodología de “clase invertida” en la que el alumno prepara la lección en casa y sólo se resuelven dudas y plantean ejercicios durante las clases presenciales. Bajo este contexto se evaluaron experiencias educativas de los cursos 0 en dos años.

A. Primera experiencia – Agosto 2012

En esta experiencia 81 estudiantes usaron la plataforma en un curso de física, aunque no todos accedieron a la plataforma. Este acceso on-line fue una actividad no obligatoria. En esta primera experiencia se evaluó mediante la combinación de los parámetros descritos en la sección III mediante un análisis retrospectivo. Para ello se usaron visualizaciones y análisis estadístico para ver resultados a nivel general de cómo había sido la interacción de los estudiantes con la plataforma así como para buscar casos específicos de estudiantes con un comportamiento que denotara casos específicos que pueden suceder en plataformas de aprendizaje.

Por ejemplo en la figura 2 se muestra una visualización en la que se estima la tasa de abandono en ejercicios de los estudiantes, mediante la comparación de la cantidad de ejercicios accedidos (círculo verde) con la cantidad de ejercicios resueltos correctamente (círculo azul). De esta manera cuanto más larga es la línea que los conecta, mayor es la tasa de abandono en ejercicios del estudiante. Se puede observar algunos estudiantes que resuelven correctamente todos los ejercicios (aparece un solo círculo verde) y otros con grandes tasas de abandono.

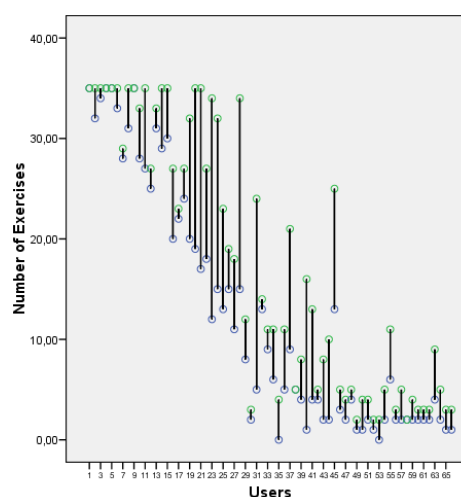


Figura 2. Estimación de la tasa de abandono por combinación de parámetros.

Otro ejemplo es la figura 3 en la que se representan los parámetros de hábitos resolviendo ejercicios de evitador de vídeos (azul), evitador de pistas (verde), usuario no reflexivo (beige) y abuso de pistas (morado), para los 19 estudiantes más activos de la experiencia. Se pueden

¹ <https://cloud.google.com/appengine/docs/python/config/cron>

² <https://cloud.google.com/appengine/docs/python/taskqueue/>

observar algunos estudiantes que tienen algunos de estos comportamientos negativos muy elevados, lo que permitiría al instructor realizar recomendaciones a estos estudiantes.

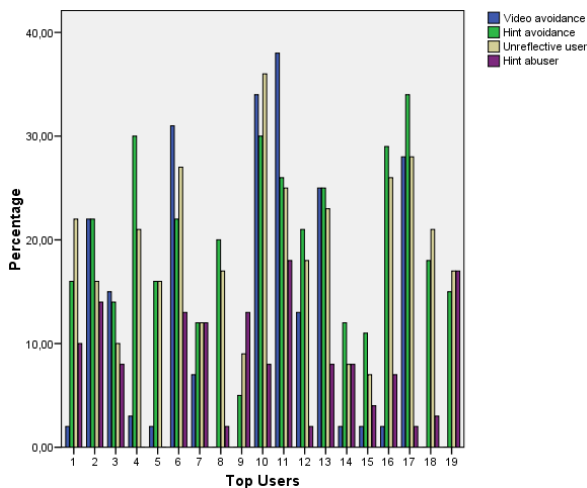


Figura 3. Parámetros de comportamiento resolviendo ejercicios de los 19 estudiantes más activos.

B. Segunda Experiencia – Agosto 2013

Después del éxito del primer año en la experiencia de los cursos 0 con Khan Academy, se repitió al año siguiente pero con mayor preparación y más cursos. En este caso, se prepararon cursos 0 de más temáticas para los cursos de física, química y matemáticas. En total el número de alumnos que interaccionaron con los cursos fueron de 167 en física, 73 en química y más de 243 en matemáticas. En esta experiencia se integró la plataforma ALAS-KA dentro de las instancias de Khan Academy de cada curso y se fue utilizando a tiempo real conforme la experiencia avanzaba. Esto implica que la herramienta de analítica de aprendizaje quedó en disposición de instructores y estudiantes para ser accedida en cualquier momento del proceso.

Se ponen dos ejemplos de visualización provenientes de ALAS-KA, en primer lugar de tipo individual en la figura 4 en la que se comparan dos estudiantes. Se puede observar los indicadores de hábitos de ludificación y elementos opcionales (primera barra) del Estudiante A y el Estudiante B, que aparecen comparados con la media de la clase (segunda barra). Ambos estudiantes emplearon una cantidad de tiempo similar en la plataforma, sin embargo como se puede ver el Estudiante A no estuvo interesado en las medallas ni en los elementos opcionales de la plataforma, mientras que el estudiante B estuvo muy interesado y se puede ver que sus parámetros son muy superiores a la media de la clase. Esta información se puede usar para adaptar el aprendizaje de cada estudiante en función de sus preferencias.

El segundo ejemplo que aparece en la figura 5 se trata de la visualización de clase para los indicadores de ejercicios y vídeos abandonados. Se puede observar como las tasas de abandono en ejercicios son mayores que en vídeos. Esto puede estar relacionado con el hecho de que resolver correctamente un ejercicio puede ser más complicado que simplemente terminar de ver un vídeo.

VI. CONCLUSIONES

En este artículo hemos conectado diversos trabajos

previos en analítica de aprendizaje para darles un marco común que ayude a entender el flujo entero que se ha

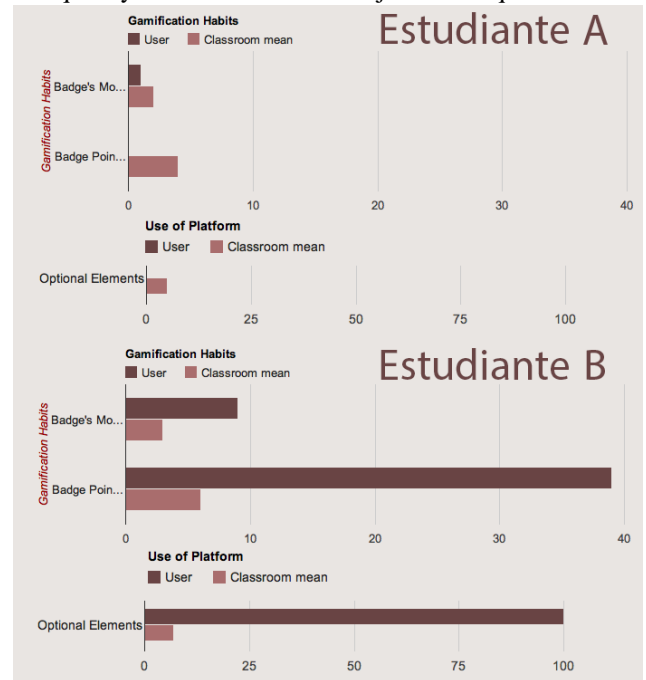


Figura 4. Comparación de los hábitos de ludificación y uso de elementos opcionales de dos estudiantes en ALAS-KA.

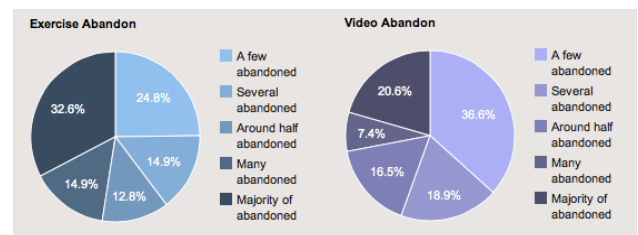


Figura 5. Porcentaje de ejercicios y vídeos abandonados de ALAS-KA.

seguido para su implementación y aplicación en un entorno usando Khan Academy. Se han explicado un conjunto de parámetros de alto nivel para la plataforma Khan Academy a partir del procesamiento de datos de bajo nivel en forma de eventos, se ha diseñado, se implementó el módulo ALAS-KA que extiende la capacidad de analítica de aprendizaje de la plataforma Khan Academy y añade nuevas visualizaciones tanto individuales como de todos los alumnos de la clase basadas en los parámetros anteriormente definidos y finalmente se ha mostrado algunos ejemplos de cómo utilizar esos parámetros y visualizaciones para evaluar el proceso de aprendizaje en experiencias reales.

El trabajo que se presenta aquí puede servir a otros investigadores en analítica de aprendizaje para hacer sus propias implementaciones y desarrollo desde cero. Durante el desarrollo del trabajo se han obtenido bastantes lecciones. Una bastante importante es que hay muchas opciones de realizar el procesado de los datos en función de cada métrica. Por otra parte, hay parámetros que pueden resultar de mucho interés, pero que no se pueden calcular por la propia semántica y funcionalidad de la plataforma.

Además, en bastantes casos la interpretación y acción a tomar en función de determinados resultados de parámetros no está clara. Por ejemplo, ¿qué hay que recomendarle a un alumno que abandona los vídeos? El que empiece a ver vídeos y no los complete puede ser porque ve el inicio del vídeo y como ya comprende las lecciones los descarta, o

porque el alumno está actuando mal en la plataforma sin completar los materiales. De momento, la decisión de recomendación y actuación se deja habitualmente al profesor para que evalúe el contexto específico para poder tomar este tipo de decisiones. Es un reto ofrecer una forma sistemática de interpretar los resultados para cada usuario, y normalmente es más efectiva una interpretación de forma individual para cada usuario en función de varios parámetros.

Otro elemento a tener en cuenta es que el procesado de datos aumenta exponencialmente conforme el curso va aumentando, ya que las cantidades de datos generadas por los usuarios son cada vez mayores. Por ello hay que tener en cuenta que esto puede provocar saturación en el servidor, o un aumento en la facturación si estamos usando algún servicio en la nube.

APÉNDICE A: CÓDIGO DE ALAS-KA, GUÍA DE INSTALACIÓN Y MANUAL DE USUARIO

Se ha realizado una página online para la aplicación desarrollada ALAS-KA, la cual se encuentra en una cuenta pública de GitHub³. Esta página contiene los siguientes contenidos:

- Código Python de ALAS-KA: Se incluye todo el código desarrollado para ALAS-KA, tanto la parte de analítica como la interfaz de visualizaciones. El código de ALAS-KA es independiente del de Khan Academy, y se debe seguir la guía de instalación para ponerlo en funcionamiento.
- Guía de instalación: Pasos a seguir para poner en funcionamiento una instancia de Khan Academy con ALAS-KA.
- Manual de usuario: Un tutorial de la funcionalidad y tipos de visualizaciones que contiene ALAS-KA, orientada tanto como para alumnos como para profesores, para que sepan que se puede hacer en la aplicación.

AGRADECIMIENTOS

Este trabajo ha sido soportado por el proyecto “eMadrid” (Gobierno Regional de Madrid) con referencia S2013/ICE-2715 y el proyecto “RESET” (Ministerio de Economía y Competitividad) con referencia TIN2014-53199-C3-1-R.

REFERENCIAS

- [1] IBM, “Informes de IBM sobre Big Data.” [Online]. Available: <http://www-01.ibm.com/software/data/bigdata/>.
- [2] “1st International Conference on Learning Analytics and Knowledge 2011.” [Online]. Available: <https://tekri.athabascau.ca/analytics/about>.
- [3] G. Siemens and P. Long, “Penetrating the Fog: Analytics in Learning and Education,” *Educ. Rev.*, 2011.
- [4] J. P. Campbell, P. B. DeBlois, and D. G. Oblinger, “Academic Analytics: A New Tool for a New Era,” *Educ. Rev.*, no. August, pp. 40–57, 2007.
- [5] D. Clow, “The learning analytics cycle: closing the loop effectively,” in *Proceedings of the 2nd International Conference on Learning Analytics and Knowledge*, 2012, pp. 134–138.
- [6] V. Alevan, B. M. McLaren, O. Roll, and K. Koedinger, “Toward Tutoring Help Seeking□; Applying Cognitive Modeling to Meta Cognitive Skills,” in *Seventh International Conference on Intelligent Tutoring Systems (ITS-2004)*, 2004, pp. 227–239.
- [7] M. Feng, N. Heffernan, and K. Koedinger, “Predicting state test scores better with intelligent tutoring systems: developing metrics to measure assistance required,” in *Proceedings of 8th International Conference, Intelligent Tutoring Systems Zhongli, Taiwan*, 2006, pp. 31–40.
- [8] H.-C. Schmitz, M. Scheffel, M. Friedrich, M. Jahn, K. Niemann, and M. Wolpers, “CAMera for PLE,” in *4th European Conference on Technology Enhanced Learning, EC-TEL*, 2009, pp. 507–520.
- [9] R. Mazza and V. Dimitrova, “Visualising student tracking data to support instructors in web-based distance education,” in *Alternate track papers & posters of the 13th international conference on World Wide Web - WWW Alt. '04*, 2004, pp. 154–161.
- [10] H. Zhang, K. Almeroth, A. Knight, M. Bulger, and R. Mayer, “Moodog: Tracking students’ online learning activities,” in *World Conference on Educational Multimedia, Hypermedia and Telecommunications*, 2007, pp. 4415–4422.
- [11] T. Tang and G. McCalla, “Smart Recommendation for an Evolving E-Learning System: Architecture and Experiment,” *Int. J. E-Learning*, vol. 4, no. 1, pp. 105–129, 2005.
- [12] P. Dwivedi and K. K. Bharadwaj, “E-Learning recommender system for a group of learners based on the unified learner profile approach,” *Expert Syst.*, vol. 32, no. 2, pp. 264–276, 2015.
- [13] P. J. Muñoz-Merino, J. A. Ruipérez Valiente, and C. D. Kloos, “Inferring higher level learning information from low level data for the Khan Academy platform,” in *Proceedings of the Third International Conference on Learning Analytics and Knowledge - LAK '13*, 2013, pp. 112–116.
- [14] J. A. Ruipérez-Valiente, P. J. Muñoz-Merino, D. Leony, and C. Delgado Kloos, “ALAS-KA: A learning analytics extension for better understanding the learning process in the Khan Academy platform,” *Comput. Human Behav.*, vol. 47, no. Learning Analytics, Educational Data Mining and data-driven Educational Decision Making, pp. 139–148, 2015.
- [15] J. A. Ruipérez-Valiente, P. J. Muñoz-Merino, C. D. Kloos, K. Niemann, and M. Scheffel, “Do optional activities matter in virtual learning environments?,” vol. 8719 LNCS, 2014.
- [16] P. J. Muñoz-Merino, J. A. Ruipérez-Valiente, C. Alario-Hoyos, M. Pérez-Sanagustín, and C. Delgado Kloos, “Precise Effectiveness Strategy for analyzing the effectiveness of students with educational resources and activities in MOOCs,” *Comput. Human Behav.*, vol. 47, pp. 108–118, 2015.
- [17] J. A. Ruipérez-Valiente, P. J. Muñoz-Merino, and C. D. Kloos, “An architecture for extending the learning analytics support in the Khan Academy framework,” in *ACM International Conference Proceeding Series*, 2013, pp. 277–284.
- [18] J. A. Ruipérez-Valiente, P. J. Muñoz-Merino, D. Leony, and C. Delgado Kloos, “ALAS-KA: A learning analytics extension for better understanding the learning process in the Khan Academy platform,” *Comput. Human Behav.*, vol. 47, pp. 139–148, 2015.

³ <https://github.com/jruiperezv/ALAS-KA>



José A. Ruipérez-Valiente completó su BSc y MSc en ingeniería de telecomunicación en la Universidad Católica de San Antonio y Universidad Carlos III de Madrid (UC3M) respectivamente. Ha trabajado tanto en el sector privado (Accenture y Grupo Multimedia Vocento) así como en el sector público (UC3M). Actualmente es estudiante de doctorado en la UC3M y asistente de investigación en el Instituto IMDEA Networks. También ha realizado una estancia de investigación de 3 meses en el MIT. Como estudiante e investigador ha recibido diversos premios que prueban sus capacidades de aprendizaje y resolución de problemas. También ha publicado en importantes revistas y conferencias internacionales relacionadas con sus áreas de investigación. Sus principales líneas en la actualidad son la analítica de aprendizaje y la minería de datos educacionales.



Pedro J. Muñoz-Merino recibió su título en Ingeniería de Telecomunicación en el año 2003 por la Universidad Politécnica de Valencia, y su Doctorado en Ingeniería Telemática en 2009 por la Universidad Carlos III de Madrid, donde en la actualidad es Profesor Visitante. Ha realizado dos estancias largas de investigación: una en Irlanda en la compañía Intel por más de 3 meses en 2005 y otra en Alemania en 2009-10 por más de 6 meses en el Fraunhofer Institute of Technology. Obtuvo la acreditación de la ANECA como Profesor Titular en Mayo de 2012. Es autor de más de 70 publicaciones científicas y ha participado en más de 20 proyectos de investigación. Ha sido miembro del comité de programa en diferentes conferencias así como invitado a impartir charlas en diferentes eventos en temas de analítica de aprendizaje y minería de datos educacionales. Es Senior Member del IEEE desde 2015.



Carlos Delgado Kloos recibió su PhD en Ciencias de Computación por la Universidad Politécnica de Munich y su MSc en Ingeniería de Telecomunicación por la Universidad Politécnica de Madrid. Desde 1996, es Catedrático del departamento en Ingeniería Telemática de la Universidad Carlos III de Madrid, donde también es director del "Máster en Gestión y Producción en e-Learning" online, director de la Cátedra UNESCO "Educación Digital Escalable para Todos" y del grupo de investigación GAST. También es Vicerrector de Estrategia y Educación Digital. Coordina la red eMadrid en tecnología educativa de la Comunidad de Madrid. Es Senior Member del IEEE. Sus principales intereses se encuentran en investigación en tecnologías educativas.