

Evaluación visual de las relaciones entre participación de los estudiantes y sus resultados en entornos de e-learning

Diego Alonso Gómez-Aguilar, Francisco J. García-Peñalvo, Roberto Therón

Title—Visual assessment of the relationships among students' participation and their outcomes on eLearning environments.

Abstract— Investigaciones recientes han confirmado el impacto del uso de los medios electrónicos en los resultados de los estudiantes. Además en anteriores trabajos se ha estudiado la relación existente entre del acceso a la lectura y publicación en un foro y el acceso a la lectura de los recursos con el desempeño de los estudiantes, sin embargo, poco se sabe de la relación de la periodicidad temporal de estas actividades con los resultados de los alumnos. Además algunos estudios han realizado diferentes categorizaciones de las actividades de los estudiantes demostrando su utilidad en la predicción y entendimiento del aprendizaje del alumno. En este estudio se ha implementado una visualización para encontrar estos aspectos mediante técnicas de analítica visual. Los resultados de este estudio mostraron que no sólo la participación en foros de discusión, como la lectura o la publicación, sino también el acceso a la lectura de los recursos, y la periodicidad con que estas actividades se realizan pueden alertar sobre el desempeño del estudiante. Esto se produjo después de comprobarse la existencia de una fuerte relación de influencia entre la frecuencia de acceso a la lectura y publicación a foros de discusión, así como entre la frecuencia de acceso a la lectura a los recursos con los resultados de los estudiantes. Sin embargo, y además se ha encontrado que la periodicidad con que estas actividades se realizan a lo largo del curso es también un factor influyente y puede analizarse visualmente, gracias a que se obtiene una mejor visualización de los patrones de rendimiento de los estudiantes.

Index Terms— e-learning, online interaction indicators, academic performance, learnin analytics, visualization, parallel coordinates.

Diego Alonso Gómez Aguilar is with the Computer Science Department /GRIAL Research Group University of Salamanca, Spain (phone: +34 923 294400; fax: +34 923 294514; e-mail: dialgoag@usal.es).

Francisco J. García-Peñalvo, is with the Computer Science Department / Science Education Research Institute / GRIAL Research Group University of Salamanca, Spain (e-mail: fgarcia@usal.es).

Roberto Therón is with the Computer Science Department of University of Salamanca, Spain (e-mail: theron@usal.es).

DOI (Digital Object Identifier) Pendiente

I. INTRODUCCIÓN

LA creciente participación de los estudiantes, la omnipresencia de las tecnologías de la información y la comunicación (del inglés *Information and Communication Technologies*, TIC), y el beneficio potencial de la observación del proceso de aprendizaje han llevado a la aparición en los últimos años de un campo de investigación en la que la tecnología se utiliza para supervisar, analizar, deducir y actuar en el desarrollo de una experiencia de aprendizaje, el “*Learning Analytics*” (LA). Una definición de esta área, y muy adecuada para el presente trabajo, es:

“*Learning analytics is the measurement, collection, analysis and reporting of data about learners and their contexts, for purposes of understanding and optimising learning and the environments in which it occurs.*”[1].

Los estudiantes interactúan con numerosos recursos (documentos, foros, chats, herramientas específicas), con los instructores, entre ellos, etc. La interacción se identifica como un componente fundamental del proceso de aprendizaje. Anderson y Garrison han identificado seis tipos de interacción que pueden tener lugar en un ambiente de aprendizaje a distancia [2]:

- 1) “*learner-teacher*”,
- 2) “*learner-content*”, y
- 3) “*teacher-content*”.

Dentro de cada uno de los componentes de las interacciones anteriores existen las siguientes interacciones adicionales:

- 4) “*learner-learner*”,
- 5) “*teacher-teacher*”, y
- 6) “*content-content*”.

Es importante mencionar que es Moore en 1989 [3] quien ha previamente identificado tres de las seis interacciones mencionadas por Anderson y Garrison: “*learner-teacher*”, “*learner-content*”, y “*learner-learner*”. Sin embargo este estudio considera que todas las interacciones tienen importancia en el proceso de entendimiento de la enseñanza y aprendizaje. La investigación sobre el aprendizaje en línea define la participación como la intervención de diferentes formas y con diferentes unidades de análisis [4-7].

La participación puede ser detectada a partir de indicadores numéricos, como el número de accesos a la plataforma, el

número de mensajes escritos, análisis de contenido de la calidad de los mensajes, la relación entre la lectura y la escritura, y la percepción que los estudiantes tienen de la actividad a través de entrevistas, cuestionarios e informes de evaluación del curso realizados por el alumno.

El presente trabajo se basa en esta idea, y aborda el problema desde un punto de vista de análisis visual, la próxima sección examina los trabajos relacionados, en la tercera se presenta una propuesta, una descripción de esta y la técnica utilizada, posteriormente se analiza un caso de estudio y, finalmente, se presentan las conclusiones del trabajo.

II. TRABAJOS RELACIONADOS

La minería de datos para la Educación (del inglés, *Educational Data Mining*, EDM) es un campo que persigue un objetivo similar al de la analítica académica (del inglés, *Academic Analytics*, AA). En EDM se propone el uso de diversas técnicas (análisis estadístico, aprendizaje automático, minería de datos, etc.) para resolver los problemas de la investigación educativa y entender el entorno en el que los estudiantes aprenden [8].

El uso de las TIC han provocado la aparición de gran cantidad de eventos registrados durante el proceso de aprendizaje. Algunos de estos conjuntos de datos están siendo recogidos, etiquetados y organizados para ser compartido dentro de la comunidad de investigación [9]. Técnicas de minería de datos que ya se utilizan en otros contextos [10], y su aplicación a contextos educativos han sido objeto de un número creciente de grupos de investigación. Las regresiones [7, 11] son un ejemplo de esto además de ser otro de los métodos usados con esta finalidad.

El término AA se volvió relevante en la comunidad de investigación después de haber sido acuñado por Goldstein y Katz en una obra fundamental que la define como la aplicación de herramientas de inteligencia de negocios, ya están en uso en los negocios y en el marketing, y ya también para el área de aprendizaje [12]. La principal aplicación de AA es ir más allá del simple informe y sugerir mejores decisiones en instituciones académicas mediante la observación y análisis grandes cantidades de eventos que ocurren en sus actividades.

El artículo caracteriza al AA como un "motor" de trabajo con cinco pasos: captura, informe, predicción, actuación y refinamiento. Los ejemplos presentados en este trabajo se refieren sobre todo al problema de la detección de "estudiantes en riesgo", es decir, aquellos alumnos que podrían abandonar un curso o abandonar sus estudios. Hay una descripción detallada de cómo la detección de estas situaciones puede beneficiar a las instituciones. Un ejemplo de la utilización de AA es el proyecto Señales [13].

A diferencia de la opinión de AA, donde se utilizan los datos capturados para tomar decisiones a nivel institucional, el objetivo de la analítica del aprendizaje (del inglés, *Learning Analytics*, LA) es el uso de estos datos y cualquier otra observación adicional que se pueden obtener, y se utiliza para tener un impacto directo sobre los estudiantes, los profesores y los detalles del proceso de aprendizaje. Por ejemplo, el análisis

de la interacción entre los estudiantes y qué provocó una inmediata participación de los estudiantes o el cambio y la adaptación del material del curso.

El aspecto que las tres áreas (LA, AA, EDM) tienen en común es la necesidad de una gran cantidad de observaciones del proceso de aprendizaje. Estas observaciones pueden provenir de diferentes fuentes, en diferentes formatos, en diferentes momentos, y necesitan ser procesadas de manera que cualquiera de los resultados útiles, requeridos por cada área, se puedan obtener [1, 14].

Por otra parte, el programa de investigación y desarrollo

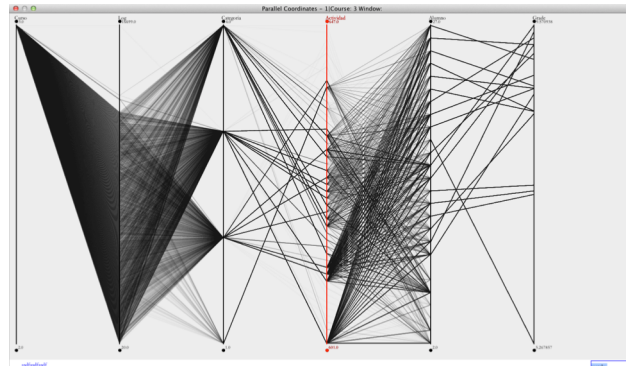


Fig. 1 Vista global de las coordenadas paralelas, donde se percibe por medio de las diferentes oscuridades de las figuras formadas por las líneas que se entrecruzan y correlacionan de eje en eje, información a primera vista. Por ejemplo se puede observar a la derecha del eje en rojo en la parte central de la imagen, cuatro bifurcaciones mas oscuras que el resto, éstas corresponden a las actividades mas realizadas ("view resource", "view assignment", "view quiz", "view course").

(<http://analiticavisual.es/images/dialogo/VGlobal.png>)

para la analítica visual [15] proporciona una lista completa de los aspectos clave que influyen en la analítica visual (del inglés, *Visual Analytics*, VA), el proceso por el cual los usuarios obtienen una idea de datos complejos. Para la discusión aquí, los aspectos más relevantes son: la ciencia del razonamiento analítico y las técnicas de interacción.

La VA utiliza las capacidades de las herramientas de visualización que ayudan a los usuarios a tomar decisiones acerca de los datos. Es importante crear herramientas de visualización que maximizan la capacidad humana para percibir y comprender información compleja y dinámica. Sin embargo, aunque las representaciones visuales proporcionan una dirección inicial para el análisis de datos, estos no serán eficaces sin mecanismos de interacción que permiten a los usuarios explorar datos [16].

El nivel de supervisión de los medios de comunicación se ha convertido recientemente en un elemento importante. La importancia de proporcionar a los profesores en "tiempo real" datos relacionados con la creación de interacciones es fundamental para informar al docente de las modificaciones necesarias de las actividades, y con esto conseguir un mejor aprendizaje en línea. Hay una fuerte evidencia que sugiere que dentro de las actividades bien estructuradas, los procesos de construcción del conocimiento alcanzan niveles más altos de pensamiento crítico y que los estudiantes son capaces de establecer y mantener grupos cohesivos. Esta prueba sirve

para justificar la necesidad de un análisis automatizado de Redes Sociales (del inglés, ARS), existen herramientas que son capaces de ofrecer “análisis en tiempo real” para los tutores y profesores [17]. Algunos estudios utilizan el número de operaciones de subida de eventos y el número de operaciones de actualización de archivos y otros eventos medibles para investigar el desempeño de los estudiantes [14]. Cabe mencionar que no existe una aplicación que permita un análisis de ambientes de aprendizaje virtual que contenga la posibilidad de explorar todo un campus completo.

Las coordenadas paralelas son una forma común de visualizar los datos de alta dimensión en el espacio limitado y permitir la exploración de las tendencias [18]. En [19] Los ejes verticales representan diferentes métricas. Las métricas difieren sobre la base de los datos disponibles. Los indicadores para el tiempo dedicado y el uso de los recursos disponibles en la iteración 1, 2 y 3 son el tiempo total pasó en el campo, el tiempo promedio de permanencia en un recurso, la cantidad de recursos utilizados y la mediana del tiempo del día cuando los estudiantes trabajan. A cada estudiante se muestra como una línea poligona.

El término AA, con el fin de evaluar el grado de participación en los cursos en línea, ha adoptado diversos enfoques metodológicos. En cuanto a los métodos cuantitativos, el acceso de los estudiantes a la zona de grupos y su acceso a las áreas de comunicación [5] y la publicación de mensajes [20, 21] se ha utilizado a menudo [6].

La participación en la actividad en términos de cantidad de mensajes publicados de los estudiantes en el foro parece estar influenciada por la calidad de las infraestructuras puestas a disposición de los estudiantes [22, 23]. Un elevado número de contribuciones de los estudiantes no siempre se correlacionan con el rendimiento de aprendizaje. De hecho, Hwang y Arbaugh [24] encontraron que los que escriben menos mensajes, pero participan en discusiones sobre una serie más amplia de temas, logra mejores resultados en la prueba de conocimiento final en comparación con los que escriben más, pero se limitan a una sola tema de discusión.

Por otra parte y no sólo tomando en cuenta las actividades en los foros, Agudo-Peregrina [11] presenta tres diferentes tipologías de interacción existentes en el e-learning y analiza la relación de sus componentes con el rendimiento académico de los estudiantes. Los tres diferentes clasificaciones se basan en los agentes implicados en el proceso de aprendizaje, la frecuencia de uso y el modo de participación, respectivamente. Los principales hallazgos de la investigación son: a) que las clasificaciones basadas en agentes ofrecen una mejor explicación del rendimiento académico de los estudiantes; b) que al menos uno de los componentes en cada tipología predice el rendimiento académico, y c) que el alumno-profesor y alumno-alumno, evaluación de los estudiantes, y las interacciones activas, respectivamente, tienen un impacto significativo en el rendimiento académico, mientras que los otros tipos de interacción no se relacionaron con el rendimiento académico significativamente. Por otro lado en [25] se corroboró mediante análisis visual que la interacción

estudiante-profesor y estudiante-estudiante es directamente proporcional al rendimiento de los estudiantes, y también se encontró que la actividad “*view resource*” es un factor influyente.

La variedad de marcos teóricos presentados a anteriormente ilustra los avances recientes de captar y estructurar patrones de aprendizaje que están surgiendo como consecuencia de la ubicuidad y accesibilidad de la información en la era digital. Todos estos tratan de reaccionar ante los recientes cambios de las principales características de los procesos y modelos de aprendizaje. Como tal, todos ellos ponen de relieve la naturaleza activa, colaborativa del aprendizaje en la era digital, la necesidad de seleccionar, procesar y analizar la información en la formación *online* de acuerdo con la relevancia y el papel central del alumno en el proceso de aprendizaje - no como un receptor pasivo de información, sino como un ente activo, autor creativo y crítico, co-creador, comentarista y evaluador. Las instituciones de educación y formación deben responder a estos cambios. La explotación de las aplicaciones de computación para propósitos educativos es y puede resultar un medio de hacerlo. Al igual que con la mayoría de las áreas de investigación, la terminología no es tan homogénea como sería deseable, pero se podría decir que los términos de EDM, AA, LA, VA y los estudios mencionados anteriormente, cubren la mayor parte de este trabajo. Tomando en cuenta lo antes mencionado, en la siguiente sección se presenta la propuesta y un caso de estudio.

III. PROPUESTA Y DESCRIPCIÓN Y TÉCNICA

Para responder, sobre la base de las cuestiones mencionadas de la sección de trabajos relacionados, si es posible encontrar una relación entre el rendimiento de los estudiantes y la participación a través de los cambios visibles en una herramienta de análisis visual y confirmar esta hipótesis, se analizaron los siguientes casos:

- 1) La relación entre la frecuencia de acceso a las publicación de los foros y la frecuencia de acceso a la lectura de las discusiones con la nota del estudiante. Fig. 2;
- 2) La relación entre la frecuencia de acceso a recursos de lectura y las calificaciones del estudiante. Fig. 6, Fig. 7 y Fig. 8.

Para este fin se ha implementado una visualización interactiva multidimensional usando la técnica de coordenadas paralelas. Esta herramienta de analítica visual aborda, como anteriormente [25-27], el análisis de los datos por medio del uso de servicios web en un entorno virtual de aprendizaje (del inglés *Virtual Learning Environment*, VLE).

Basándose en [25] se tomó ventaja de sus beneficios, es decir, la posibilidad de utilizar en cualquier momento dado, que es independiente de la versión del ambiente virtual de aprendizaje, además de poder ser utilizada *offline*, *online*, en web o aplicación de escritorio, etc. Este desarrollo amplía las posibilidades de la herramienta de analítica visual con el fin de revelar otro tipo de patrones. Esta herramienta esta dirigida

tanto a estudiantes, como profesores y académicos de una institución educativa; ya que los datos a los que el usuario puede acceder dependen de los permisos que tenga en el VLE, estos pueden variar desde un curso hasta un campus universitario completo; teniendo la capacidad de exploración y comparación de estos. Por motivos de espacio las imágenes presentadas en este estudio se han elegido ya que son las que mejor representan el patrón encontrado.

Desde la introducción temprana de Inselberg [28], las coordenadas paralelas han sido ampliamente utilizadas para visualizar datos multivariados [29]. Las coordenadas paralelas son una de las más famosas técnicas de visualización, y entre los temas más comunes de trabajos académicos en la visualización. Aunque en un principio confuso, es una herramienta muy poderosa para la comprensión de conjuntos de datos numéricos multidimensionales.

Las coordenadas paralelas [28] han sido ampliamente utilizadas para el análisis de conjuntos dibujando dimensiones como ejes paralelos y datos del hiperespacio con poli-líneas que conectan sus valores escalares en los ejes, las coordenadas paralelas pueden representar los datos de N-dimensiones en un espacio de 2 dimensiones. Sin embargo, cuando los tamaños de los conjuntos de datos se vuelven tan grandes que millones de líneas de llenan a la pantalla, estas pueden llegar a ser demasiado densas para ser interpretadas.

La forma habitual de describir las coordenadas paralelas sería hablar de espacios de alta dimensión representados con la técnica de establecer los ejes de coordenadas en paralelo en lugar de ortogonales entre sí. Para que sea más fácil de interpretar y entender, imagine una tabla de datos en donde cada una de estas columnas están asignadas a un eje vertical en la imagen de la Fig. 1. Cada valor de datos le corresponde una posición a lo largo de la línea, dependiendo del valor del dato terminaría entre el mínimo en la parte inferior y el máximo en la parte superior. Una colección pura de puntos en estas posiciones de los datos no sería extremadamente útil, por lo que los puntos que pertenecen a un mismo registro (fila) se conectan con líneas. Para dar un poco de sentido a esto, la forma más fácil es olvidar la parte de los registros que abarcan todo el ancho, y mirar el espacio entre cada par de ejes. De esta forma se puede encontrar los patrones en cada uno de los espacios entre los ejes.

Además de un poco de experiencia en la lectura de coordenadas paralelas, la mejor manera de conocer un conjunto de datos es utilizando las técnicas de la interacción que proporcione la visualización. La principal de ellas en coordenadas paralelas se llama *Brushing*. Esta es una técnica muy efectiva para especificar un enfoque explícito en visualización de información [30, 31]. En esta el usuario marca activamente sub-conjuntos de los datos de ajuste mediante el uso del ratón, seleccionando zonas de intersección con las líneas, dibujadas como rectángulos de selección en muchos casos o rangos en los ejes (mire los pequeños rectángulos sombreados en las imágenes de selección de subconjuntos, por ejemplo de la imagen 3 en adelante). Si se utiliza en conjunción con múltiples puntos de vista vinculados,

el *Brushing* (Fig. 2 en adelante) puede permitir a los usuarios entender las correlaciones a través de múltiples dimensiones [30, 32, 33]. Cabe mencionar que la visualización implementada permite la multi-selección de subconjuntos en los ejes verticales o fuera de ellos. Otra técnica es el uso de las diferentes agrupaciones de las actividades, para colorear las líneas de las coordenadas paralelas con el objetivo de identificar patrones.

El uso de categorías tiene importancia para este estudio dado que se requirió implementar la manera de categorizar dinámicamente las actividades de los estudios mencionados en la introducción. Además de permitir el uso de categorías dentro de los ejes de las coordenadas paralelas y la coloración de las líneas en base a estas, también permite la creación de agrupaciones personalizadas y generar un archivo con estas para su importación y exportación. Finalmente, los métodos de interacción como: el filtro de búsqueda a través de la caja de texto abajo a la izquierda de la visualización, el ocultamiento y reordenación de columnas también son de vital importancia para un análisis visual.

¿Por qué una herramienta visual en vez de un análisis puramente numérico? Por un lado, por medio de un análisis numérico, se tiene una pregunta específica a responder siendo este un análisis poco dinámico y por otro lado no se tiene una pregunta específica sino que la que es la exploración y satisfacción de la curiosidad la que lleva a obtención de nuevo conocimiento y nuevas preguntas a responder. Para este segundo caso, las herramientas basadas en técnicas de analítica visual explotan las capacidades visuales del ser humano logrando la rápida comprensión de las correlaciones existentes o encontradas en los datos, la comparación de elementos de los datos y la exploración libre a través de las herramientas visuales. Por este motivo se realizó y recomienda la implementación y uso de herramientas de analítica visual.

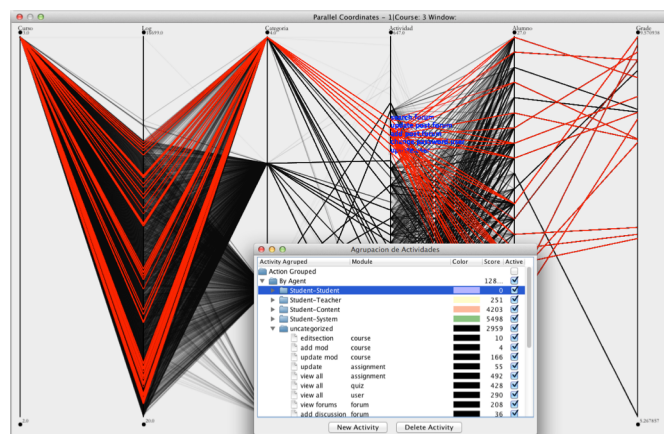


Fig. 2 Vista del subconjunto de las actividades comunes con la interacción en foros como: *search forum*, *update post forum*, *add post forum*, etc.

(<http://analiticavisual.es/images/dialogoag/StudentGoodGrade.png>)

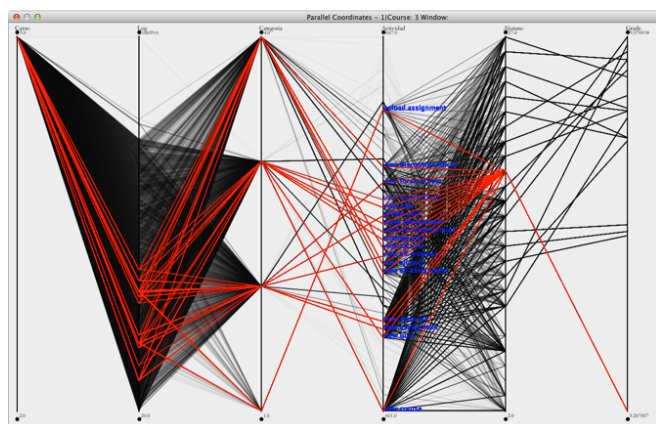


Fig. 3 Vista del subconjunto de las actividades realizadas durante todo el curso de un alumno con una nota de 5.2.

(<http://analiticavisual.es/images/dialogoag/StudentLow.png>)

IV. CASO DE ESTUDIO

Para el caso de estudio que se presenta se ha utilizado la herramienta de análisis visual que implementa la técnica de las Coordenadas Paralelas antes mencionada. Los datos utilizados se obtuvieron de una plataforma Moodle con el fin de analizar un curso. La duración de este es de Noviembre 2011 a Mayo 2012, casi el ciclo escolar con 13000 registros de actividades que fueron realizados por un total de 30 estudiantes.

El análisis consiste en la selección, la unificación, la normalización y el filtro de los eventos del registro del VLE. Ese filtro se hizo para eliminar los mensajes del foro que no son relevantes (como "sonrisas", "abrazos", "hola", etc.) [34].

Los datos para este estudio se recogieron durante todo el curso. En primer lugar, el número de accesos, la lectura de acceso y mensajes para cada discusión, se contó el número de accesos de lectura a un recurso para producir datos cuantitativos para el análisis, se obtuvo el modulo de Moodle utilizado y el contexto con el fin de realizar la categorización adecuadamente.

Los datos consisten en las filas: "Log", "Curso", "Alumno", "Rol", "Actividad", "Modulo", "Categoría", "Grade" como puede verse en cualquiera de las figuras del artículo.

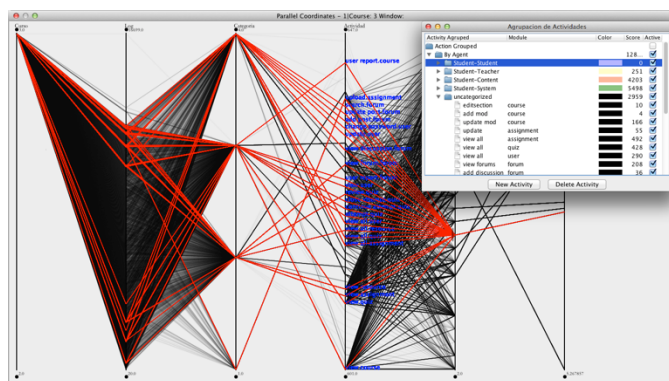


Fig. 4 Vista del subconjunto de las actividades realizadas durante todo el curso de un alumno con una nota de 6.5.

(<http://analiticavisual.es/images/dialogoag/StudentMed.png>)

No obstante es posible cambiar el orden y ocultar o mostrar cualquiera de estas. Para este estudio se ha ocultado la columna de "rol" y "Modulo" además se ordenaron los ejes de la siguiente manera: "Curso", "Log", "Categoría", "Actividad", "Alumno", "Grade" (para mejor claridad vea la Fig. 1, la cual no tiene seleccionado ningún subconjunto de los datos). A primera vista en las imágenes se puede reconocer fácilmente unos triángulos formados por las líneas entre un eje vertical y otro, que también pueden apreciarse como cilindros Fig. 1. El triángulo más grande es el que se reconoce entre el primer y segundo eje (eje de cursos y de log) esto es dado que solo tiene un curso (primer eje).

Entre el eje del log, que representa todas las actividades realizadas en orden cronológico de abajo hacia arriba, y el tercer eje, el número de figuras triangulares formadas solo puede ser un número entero, ya que es el de las categorías y no hay más de cuatro, por lo que todas las líneas tienen que pasar por un pequeño número de puntos. Además, en el espacio entre Log y Categoría, se puede decir que la categoría que tiene menos actividades realizadas es la 1, la categoría 3 es la más recurrida seguida por la 2 y después la 4.

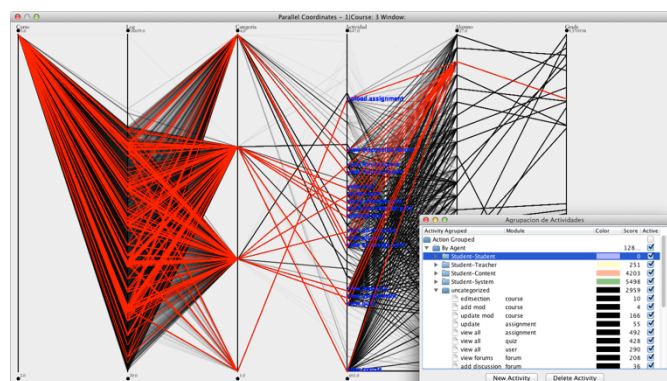


Fig. 5 Vista del subconjunto de las actividades realizadas durante todo el curso de un alumno con una nota de mas de 8.

(<http://analiticavisual.es/images/dialogoag/StudentHight.png>)

Esto como resultado de observar cuál es la más oscura de las figuras formadas entre los ejes, y resultado de usar la transparencia de las líneas, siendo la figura formada más oscura la de mayor frecuencia. Para algunos casos sólo hay que seguir las líneas y ver cómo se cruzan: un conjunto grande de líneas que se cruzan son un indicio de una relación inversa.

La correlación es mucho más directa en algunos conjuntos de datos, pero para este caso de estudio y con el objetivo de corroborar las correlaciones de los estudios antes mencionados, como por ejemplo:

- cuanto mayor interacción mejor será nota,
- cuanto mayor sea el número de entradas de lectura a los foros y publicación en estos mejor será al nota,
- cuanto mayor es la frecuencia de lectura de recursos y la participación en los foros mayor será la nota.

Para comprobar estas hipótesis se exploraron los datos seleccionando diferentes subconjunto de datos, y se han elegido las imágenes más representativas por motivos de espacio. Si se observan las figuras Fig. 3, Fig. 4 y Fig. 5, en ellas se seleccionó un alumno con nota baja, media y alta respectivamente con el objetivo de comprobar la hipótesis a). Se puede observar, a primera vista, que a mayor frecuencia de actividades (nótese la cantidad de líneas en rojo que cruzan el segundo eje de izquierda a derecha, eje de *Log*) es mas alta la nota. Sin embargo entre el alumno de nota media (Fig. 4) y nota baja (Fig. 3), la diferencia es muy pequeña lo que hace resaltar el patrón de periodicidad. Es decir que no solo la cantidad de actividades que realice el estudiante es representativo para deducir que obtendrá buena nota, sino también la periodicidad con la que se realizan estas. Para el caso de el alumno de baja nota (Fig. 3) solo realizó actividades al inicio del curso (mírese las líneas rojas que cruzan el eje de *Log* que solo se colocan en la parte inferior); en comparación con el alumno de nota media que en la Fig. 4 sus actividades fueron realizadas tanto al inicio del curso como al final.

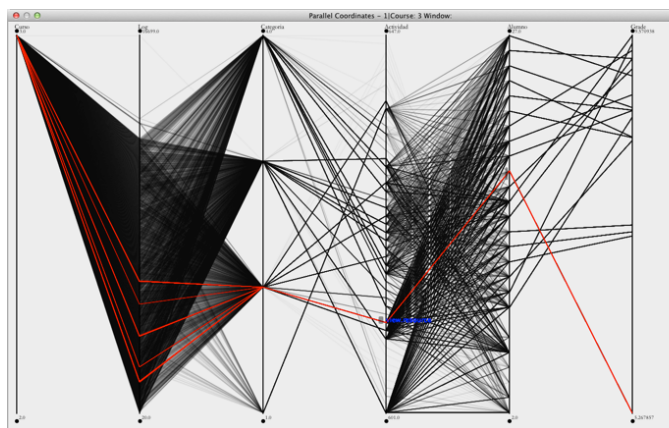


Fig. 6 Vista del subconjunto de las actividades de “view resource” realizadas durante todo el curso de un alumno con una nota de 5.2.

(<http://analiticavizual.es/images/dialogoag/StudentLow-R.png>)

Además se puede observar que no son las mismas actividades las que se realizan, razón por la cual se ha realizado el siguiente experimento: se seleccionó el conjunto de categorías de actividades y subconjuntos de las actividades de estas que fueron estudiadas anteriormente en el trabajo mencionado [11] con el fin de comprobar esa relación en este conjunto de datos y develar nuevos patrones encontrando los siguientes resultados.

En la categoría 4 mencionada en el estudio como representativa se denotó (ver Fig. 2) que las actividades “search forum, update post forum, add post forum” fueron realizadas únicamente por los alumnos con nota buena y alta. Con esto se puede comprobar la relación de que cuanto más sea la frecuencia de las actividades de esta categoría mayor será la nota y en específico de estas actividades.

Posteriormente, con el objetivo de comprobar la hipótesis c) en este conjunto de datos y observar cómo se manifiesta la periodicidad, Fig. 6, Fig. 7 y Fig. 8 muestran la selección del subconjunto de la actividad “view resource” de un estudiante

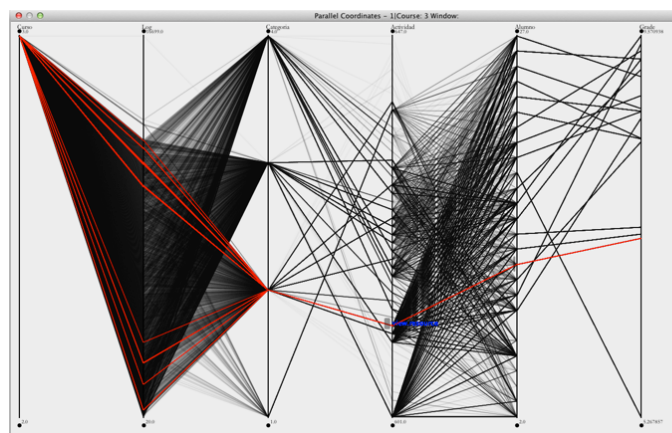


Fig. 7 Vista del subconjunto de las actividades de “view resource” realizadas durante todo el curso de un alumno con una nota de 6.5.

(<http://analiticavizual.es/images/dialogoag/StudentMedium-R.png>)

con nota baja, media y alta respectivamente. En estas imágenes claramente se observa que se repite el patrón, esto es, que cuanto más frecuencia de actividad mejor será la nota. Sin embargo no siempre significa mejor nota cuanto mayor frecuencia de actividad, pero la tendencia general es claramente esa. Cabe mencionar que también el patrón de la periodicidad se ve en este ejemplo, ya que en las imágenes Fig. 6 y Fig. 7 se puede ver que la diferencia de frecuencias no es muy grande pero sin embargo la periodicidad de la Fig. 7 está mejor distribuida (entre el inicio y el final del curso) en comparación con la Fig. 8, teniendo un efecto positivo en la nota, que varía de 5.2 a 8.

Finalmente, y con el objetivo de mostrar la potencialidad de esta técnica en comparación con otras herramientas y representaciones, y mostrándola como ejemplo de la posible exploración, con el uso de estas técnicas, de grandes cantidades de datos educativos, tal como se necesita en AA, puede observar se la Fig. 9.

Antes de comentar alguna información que a simple vista se denota, es preciso comentar que estos datos están tomados de un campus completo. Este campus universitario comprende

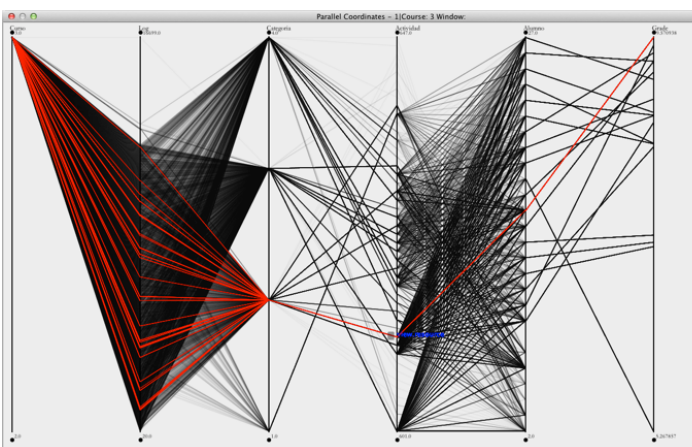


Fig. 8 Vista del subconjunto de las actividades de “view resource” realizadas durante todo el curso de un alumno con una nota de 9.6.

(<http://analiticavizual.es/images/dialogoag/StudentHightl-R.png>)

diez meses de junio del 2008 a abril del 2009 con un total de más de 9 millones de registros, mas de 2300 cursos, mas de 4800 foros, cerca de 12000 discusiones, casi 35 mil publicaciones en foros y cerca de 20 mil estudiantes. La imagen muestra la estructura de todos estos cursos y para efectos de este ejemplo se han ocultado algunos ejes de métricas e información privada, información útil para trabajos futuros de AA o DME. Sin embargo, siendo las columnas mostradas curso, foro, discusión, post, usuario y usuario que contesta a una publicación determinada, ya a simple vista se pueden observar algunos detalles que llaman la atención.

Por ejemplo, entre el segundo y tercer eje (ejes de foro y discusión), se notan unas zonas más oscuras que otras, estas develaciones serían un buen punto de partida para la exploración. Además en la columna de usuario y usuario que contesta a una publicación de foro determinado (ultimo y penúltimo eje) también se notan unas zonas más oscurecidas que otras, estos son usuarios que participan mucho más que el resto, e inmediatamente surgen nuevas preguntas como: ¿estos usuarios participan tal cantidad, que sobresalen del resto, en uno o varios cursos?, ¿los cursos son de un área en específico?, y ¿los usuarios que contestan significativamente mas a las publicaciones en foros, son profesores o estudiantes?, ¿coinciden con los que publican significativamente más que el resto de usuarios?, etc.

Si bien este es un ejemplo muy simple, muestra las estructuras típicas que se puede encontrar en la mayoría de los conjuntos de datos de este tipo.

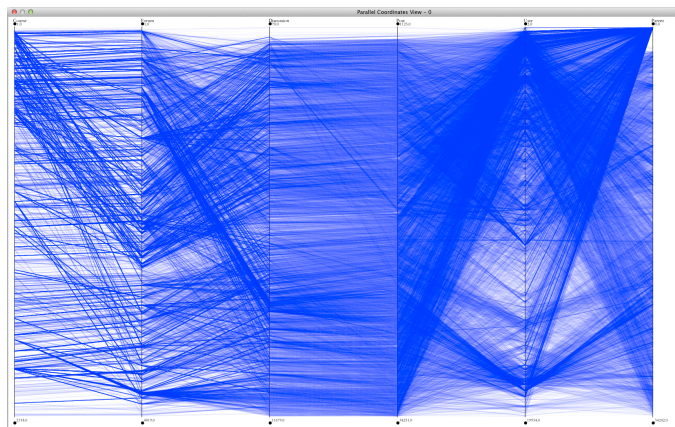


Fig. 9 Vista Global del campus completo de una universidad. Nótese con la ayuda de las transparencias y los diferentes tonos de oscuridad las zonas donde se acumula mayor frecuencia ya sea de foros de un curso o de post de un usuario específico o de discusiones de un foro.

(<http://analiticavisual.es/images/dialogoag/campusView.png>)

V. CONCLUSIONES

Como conclusión, esta visualización ayuda a revelar una tendencia con facilidad y de forma interactiva, por lo tanto proporcionar rápidamente información oculta.

Está claro que la tendencia futura será hacia una mayor adopción de aprendizaje en línea. Por lo tanto, habrá una mayor oportunidad para que profesores y administradores requieran supervisar la actividad de los estudiantes y la interacción con el contenido del curso y sus compañeros. Los análisis de estos conjuntos de datos están directamente relacionados con la participación del estudiante, así como la evaluación de las actividades de aprendizaje implementadas. Las preguntas relacionadas con cómo y cuándo los estudiantes están comprometidos, y qué actividades promueven la participación del estudiante, se puede responder mediante el seguimiento y análisis del comportamiento del estudiante en línea. Si bien los datos examinados en este trabajo indican en esta etapa, y la interpretación de los resultados discutidos aquí pueden verse influidas en cierta medida por una serie de variables exógenas, los resultados, sin embargo, ofrecen nuevas perspectivas de aprendizaje de los estudiantes que se complementan. En este sentido, el estudio proporciona una plataforma para investigar en nuevas suites de herramientas de diagnóstico que pueden, a su vez, ofrecer nuevas oportunidades y nuevos conjuntos de datos para informar a la reflexión instructor para la mejora continua y progresiva de la práctica pedagógica.

En este artículo se ha utilizado la representación en Coordenadas Paralelas de un sistema de analítica visual que es capaz de describir las interacciones en entornos en línea (sobre todo en VLE). Una primera contribución de este trabajo es demostrar, con el experimento, que existe un patrón de comportamiento entre este tipo de participación y las calificaciones del estudiante además de cómo su periodicidad afecta.

Esta visualización aporta nuevas posibilidades para el aprendizaje electrónico como una herramienta capaz de no solo ayudar al profesor por medio del reagrupamiento de actividades y que ilustra el grado de participación de los estudiantes, identifica a los estudiantes que sobresalen en la interacción con la herramienta fácilmente y encuentra las relaciones implícitas entre los participantes en los foros, pero también ayudará a encontrar patrones ocultos.

En muchos sentidos, los gráficos son la solución obvia para tratar de representar visualmente una estructura de red. Sin embargo, dadas las limitaciones de la resolución y el tamaño de la pantalla de las computadoras, los gráficos de redes sociales suelen acabar desordenado y bastante ilegible. Como contribución de este artículo se sugiere que los gráficos deben ser diseñados con técnicas como esta y con la ayuda de la interacción y algunas otras técnicas más el mantra: "*Analyze First -Show the Important -Zoom, Filter and Analyze Further -Details on Demand*" [35, 36].

AGRADECIMIENTOS

Los autores agradecen el apoyo del Ministerio de Ciencia e Innovación (proyectos FFI2010-16234 y TIN2010-21695-C02-01) y de la Consejería de Educación de la Junta de Castilla y León (proyecto SA294A12-2).

REFERENCIAS

- [1] H. Fournier, R. Kop, and H. Sitlia, "The value of learning analytics to networked learning on a personal learning environment," in *Proceedings of the 1st International Conference on Learning Analytics and Knowledge*, Banff, Alberta, Canada, 2011, pp. 104-109.
- [2] T. Anderson, and D. R. Garrison, "Learning in a Networked World: New Roles and Responsibilities," <http://hdl.handle.net/2149/801>, 1998].
- [3] M. G. Moore, "Editorial: Three types of interaction," *American Journal of Distance Education*, vol. 3, no. 2, pp. 1-7, 1989/01/01, 1989.
- [4] S. Hrastinski, "What is online learner participation? A literature review," *Computers & Education*, vol. 51, no. 4, pp. 1755-1765, 12//, 2008.
- [5] J. Davies, and M. Graff, "Performance in e-learning: Online participation and student grades," *British Journal of Educational Technology*, vol. 36, no. 4, pp. 657-663, //, 2005.
- [6] J. Kim, "Influence of group size on students' participation in online discussion forums," *Computers & Education*, vol. 62, no. 0, pp. 123-129, 3//, 2013.
- [7] F. Pascual-Miguel et al., "A characterisation of passive and active interactions and their influence on students; achievement using Moodle LMS logs," *International Journal of Technology Enhanced Learning*, vol. 3, no. 4, pp. 403-414, 01/01/, 2011.
- [8] R. S. J. d. Baker, and K. Yacef, "The State of Educational Data Mining in 2009: A Review and Future Visions," *Journal of Educational Data Mining*, vol. 1, no. 1, 2009.
- [9] K. R. Koedinger et al., "An open repository and analysis tools for fine-grained, longitudinal learner data," *1st International Conference on Educational Data Mining, EDM 2008*, pp. 157-166, 2008.
- [10] U. Fayyad, G. Piatetsky-Shapiro, and P. Smyth, "From data mining to knowledge discovery: an overview," *From Data Mining to Knowledge Discovery in Databases*, vol. 17, no. 3, pp. 37-54, //, 1996.
- [11] A. F. Agudo-Peregrina, A. Hernandez-Garcia, and S. Iglesias-Pradas, "Predicting academic performance with learning analytics in virtual learning environments: A comparative study of three interaction classifications." pp. 1-6.
- [12] P. J. Goldstein, and R. N. Katz, *Academic analytics: The uses of management information and technology in higher education*, vol. 8, Tech. Rep. December EDUCAUSE Center for Applied Research (2005), 2005.
- [13] Z. Tanes et al., "Using Signals for appropriate feedback: Perceptions and practices," *Computers & Education*, vol. 57, no. 4, pp. 2414-2422, 12//, 2011.
- [14] V.-A. Romero-Zaldivar et al., "Monitoring student progress using virtual appliances: A case study," *Computers & Education*, vol. 58, no. 4, pp. 1058-1067, 5//, 2012.
- [15] J. J. Thomas, and K. A. Cook, "The Science of Analytical Reasoning," *Illuminating the Path: The Research and Development Agenda for Visual Analytics*, pp. 32-68: National Visualization and Analytics Ctr, 2005.
- [16] A. van Barneveld, E. A. Kimberly, and P. C. John, "Analytics in Higher Education: Establishing a Common Language," *EDUCAUSE Learning Initiative*, vol. ELI Paper 1, Jan., 2012.
- [17] A. Bakharia, and S. Dawson, "SNAPP: a bird's-eye view of temporal participant interaction," in *Proceedings of the 1st International Conference on Learning Analytics and Knowledge*, Banff, Alberta, Canada, 2011, pp. 168-173.
- [18] H. Siirtola et al., "Visual Perception of Parallel Coordinate Visualizations." pp. 3-9.
- [19] S. Govaerts et al., "The student activity meter for awareness and self-reflection," in *CHI '12 Extended Abstracts on Human Factors in Computing Systems*, Austin, Texas, USA, 2012, pp. 869-884.
- [20] L. Lipponen et al., "Patterns of participation and discourse in elementary students' computer-supported collaborative learning," *Learning and Instruction*, vol. 13, no. 5, pp. 487-509, //, 2003.
- [21] R. S. Weisskirch, and S. S. Milburn, "Virtual discussion: Understanding college students' electronic bulletin board use," *Internet and Higher Education*, vol. 6, no. 3, pp. 215-225, //, 2003.
- [22] S. Cacciamani, and M. Luciani, *Una knowledge building community nella formazione universitaria on line*: Polimettrica, 2007.
- [23] D. Cesareni, and F. Martini, "Costruire conoscenza in un forum universitario," *Rassegna di Psicologia*, vol. 22, pp. 89-112, //, 2005.
- [24] A. Hwang, and J. B. Arbaugh, "Seeking feedback in blended learning: competitive versus cooperative student attitudes and their links to learning outcome," *Journal of Computer Assisted Learning*, vol. 25, no. 3, pp. 280-293, 2009.
- [25] D. A. Gómez-Aguilar, R. Therón, and F. J. García Peñalvo, "Reveal the relationships among student's participation and its outcomes on eLearning environments: Case study," *Advanced Learning Technologies, 2013. ICALT 2013. 13th IEEE International Conference on*, in press., 2013.
- [26] D. A. Gómez Aguilar et al., "Retrieval information model for Moodle data visualization," *Advanced Learning Technologies 2010. ICALT 2010. 10th IEEE International Conference on*, vol. Product Number E4055, pp. 526-527, July 5-7, 2010.
- [27] D. A. Gómez-Aguilar, R. Therón, and F. J. García Peñalvo, "Semantic spiral timelines used as support for e-learning," *Jornal of Universal Computer Science (j-jucs)*, vol. 15, no. 7, pp. 1526-1545, April, 2009.
- [28] A. Inselberg, and B. Dimsdale, "Parallel coordinates: a tool for visualizing multi-dimensional geometry." pp. 361-378.
- [29] M. Tory, S. Potts, and T. Moller, "A parallel coordinates style interface for exploratory volume visualization," *Visualization and Computer Graphics, IEEE Transactions on*, vol. 11, no. 1, pp. 71-80, 2005.
- [30] R. A. Becker, and W. S. Cleveland, "Brushing scatterplots," *Technometrics*, vol. 29, no. 2, pp. 127-142, 1987.
- [31] M. O. Ward, "XmdvTool: integrating multiple methods for visualizing multivariate data." pp. 326-333.
- [32] H. Hauser, F. Ledermann, and H. Doleisch, "Angular brushing of extended parallel coordinates." pp. 127-130.
- [33] D. L. Gresh et al., "WEAVE: a system for visually linking 3-D and statistical visualizations applied to cardiac simulation and measurement data." pp. 489-492.
- [34] C. Romero, S. Ventura, and E. García, "Data mining in course management systems: Moodle case study and tutorial," *Computers & Education*, vol. 51, no. 1, pp. 368-384, 8//, 2008.
- [35] D. A. Keim et al., "Challenges in Visual Data Analysis," in *Tenth International Conference on Information Visualization. IV' 06.*, London, England, 2006, pp. 9-16.
- [36] B. Shneiderman, "The eyes have it: a task by data type taxonomy for information visualizations." pp. 336--343.