

Algoritmo de Cómputo del Valor de las Sesiones de Usuarios para Evaluación de los Logros de un Sitio Web

Ernestina Menasalvas

Facultad de Informática UPM. Madrid España
emenasalvas@fi.upm.es

Socorro Millán

Facultad de Ingeniería. Universidad del Valle. Cali. Colombia
millan@eisc.univalle.edu.co

Esther Hochsztain

Facultad de Ingeniería - Universidad ORT Uruguay
esthoc@adinet.com.uy

Abstract

Web Mining methodology needs to adapt itself to continuous changes of web sites goals. One of the reasons of failure in web mining is that most of the web mining firms concentrate on the analysis exclusively of clickstream data. Nevertheless, this data has to be enhanced with information about the business if useful patterns want to be extracted. Today unless you can demonstrate profits you are not likely to make it or survive.

We give an algorithm to determine the user's session value. It adapts itself both to web site goals, and to changes in the criteria of the analyst of the web sequences. Our proposal is an innovative solution because it allows us to differently evaluate paths from different users and adapt the analysis to the goals of different users integrating the information from the web with the business goals.

The input of the algorithm is a values matrix that represents the value of each link that a user can traverse in the web site. We propose a method to obtain them. Experimental evaluation of 2400 sessions with four matrixes is presented.

Key words: web mining, graph based analysis, navigator session value

Resumen

La exitosa aplicación de técnicas de minería de datos en la Web requiere que éstas se adapten a los cambios continuos en los objetivos de los sitios web. Una de las razones por las cuales ha fallado la aplicación de técnicas de descubrimiento de conocimiento en datos extraídos de la web es que, en la mayoría de los casos, el análisis se ha concentrado exclusivamente en análisis de páginas y caminos más visitados sin tener en cuenta los objetivos del sitio web. Sin embargo, si se quieren extraer patrones útiles e interesantes, los datos de la web se deberían enriquecer con información relacionada con el negocio.

En este artículo se propone un algoritmo para determinar el valor de una sesión de un usuario en la web. Dicho algoritmo, tiene en cuenta las metas del sitio web, el comportamiento y perfil del usuario y los cambios en las políticas y objetivos marcados por los administradores del sitio web. La solución que se propone es innovadora en el sentido en que permite, tener en cuenta puntos de vista de distintos usuarios, e integrar la información del sitio web con las metas del negocio.

La entrada del algoritmo es una matriz de valores en la que cada casilla representa el valor que tiene avanzar desde una determinada página a otra. El artículo presenta también resultados experimentales basados en 2400 sesiones analizadas atendiendo a cuatro diferentes matrices.

Palabras Claves: Minería de datos en la web, análisis basado en grafos, valoración de sesiones de usuario

1. Introducción

El continuo crecimiento del World Wide Web, unido al entorno competitivo en el cual se mueven las organizaciones modernas, ha hecho necesario diseñar los sitios web teniendo en cuenta, como aspecto fundamental, el conocimiento que se puede extraer de las navegaciones de los usuarios que lo utilizan. Una de las formas de conocimiento más frecuentemente utilizadas consiste en descubrir cuáles son los caminos de usuario más frecuentes. Sin embargo, esto no es suficiente, haciéndose necesario integrar, por ejemplo, minería de datos con los objetivos del sitio web, con el propósito de conseguir que cada sitio web sea el más atractivo y como consecuencia el más competitivo.

La mayoría de las organizaciones que exploran el comportamiento de sus usuarios en la web utilizan, exclusivamente, datos de las secuencias visitadas ("clickstream"). Hasta el momento, uno de los principales problemas en la aplicación de técnicas de data mining en datos de la web tiene que ver con la etapa de preprocesamiento de datos. Los servidores web registran, comúnmente, una entrada por cada acceso en el archivo *log*. Entre los datos que se recogen se incluye la dirección IP, el tiempo de acceso, el método pedido, el URL de la página solicitada el protocolo de transmisión, un código de retorno y el número de *bytes* transmitido. El servidor *log* contiene, no obstante, muchas entradas que son irrelevantes o redundantes para la tarea de minería y que se requiere limpiar antes del preprocesamiento. Después de la limpieza, es necesario identificar y agrupar los datos en sesiones significativas [12]

Las técnicas inteligentes de web mining (intelligent web mining) pueden aprovechar los datos del clickstream una vez preprocesados para extraer conocimiento relacionado con la interacción de los usuarios con la Web [1][2], que se puede utilizar para tomar decisiones críticas de negocio. Sin embargo, estos datos se deben enriquecer con información relativa al negocio si lo que se espera es ofrecer a las organizaciones conocimiento interesante y útil sobre el mismo y sobre sus clientes de forma que les permita competir. De acuerdo con [3] hoy en día, a menos que se pueda obtener y demostrar ganancia, no se podrá sobrevivir.

En este sentido, en este artículo se propone un algoritmo que a la vez que tiene en cuenta la información registrada en el servidor *log* mejora el análisis tradicional, puesto que integra información del negocio. El enfoque propuesto tiene en cuenta, para el cálculo de los valores de un enlace, los datos almacenados en el archivo *log* del servidor, los objetivos del negocio y el conocimiento disponible sobre el área o contexto del negocio.

El algoritmo permite calcular los valores acumulados, durante una sesión, teniendo en cuenta, tanto el análisis del comportamiento de los usuarios como las metas cambiantes del negocio. La idea básica subyacente al algoritmo es muy similar al proceso de corrección de una prueba de evaluación de los estudiantes. En el caso de los exámenes, dependiendo de sus respuestas los alumnos suman o restan puntos a su calificación. Haciendo una analogía, las páginas visitadas por un visitante lo pueden alejar o acercar a la meta propuesta por la organización. Cuando éste se acerca a la meta, mientras visita las páginas, se añaden puntos; cuando se aleja se restan.

La solución que se propone en este artículo es innovadora porque considera diferentes caminos de evaluación a partir del punto de vista de diferentes usuarios integrando la información proveniente de la web con los objetivos del negocio. De esta manera, se ofrece un marco conceptual para analizar la evolución de las sesiones asignándoles un valor. El enfoque propuesto facilita también la detección de patrones de evolución a partir de sesiones de diferente valor.

El enfoque de representación utilizado se basa en un grafo dirigido como el propuesto en [4] y [5] y en las páginas web adaptativas propuestas en [6] [7] y [8]. El valor del cambio en la conducta de los usuarios es útil para tomar decisiones sobre la necesidad de adaptar las páginas web y sobre cómo hacerlo. Por otra parte, ésta propuesta se relaciona también con el descubrimiento de secuencias propuesto en [9] [10]. El algoritmo que proponemos, requiere también de una fase de preparación exhaustiva de los datos para identificar sesiones y usuarios tal y como se propone en [11].

El resto del artículo está organizado como sigue: En la sección 2, se presentan los conceptos básicos relacionados con el enfoque propuesto. En la sección 3, se describe el algoritmo para calcular la evolución del valor de las sesiones. En la sección 4, se presentan algunos criterios para analizar el valor de sesiones junto con un ejemplo de aplicación. La sección 5 presenta la metodología de determinación de matrices de valoración. La sección 6 presenta los resultados experimentales obtenidos al aplicar el algoritmo sobre un conjunto de 2400 diferentes sesiones. Finalmente, en la sección 7 se presentan las conclusiones y las futuras tareas de investigación que surgen del enfoque propuesto

2. Conceptos básicos

En esta sección se presentan algunos de los conceptos básicos en los cuales se apoya el algoritmo propuesto:

Sitio Web: Como en [13] un sitio web se define como un conjunto finito de páginas web.

Sea W un sitio web y sea \mathbf{W} un conjunto finito representando las páginas contenidas en W . Cada página tiene asignado un identificador único, de manera que un sitio web consistiendo de m páginas se representa como $\mathbf{W} = \{a_1, \dots, a_m\}$. $W(i)$ representa el i -ésimo elemento o página.

Dos páginas especiales, que se denotan como a_0 y a_n , corresponden a la página desde la cual el usuario entra al sitio web y la que visita antes de salir de la sesión, respectivamente [14]

Representación de un sitio web: un sitio web se considera un grafo dirigido, definido como (N, E) , donde N es un conjunto de nodos y E es un conjunto de arcos. Un nodo se corresponde con una página web y un arco con un enlace.

Páginas objetivo: Las páginas objetivo son los nodos que se desean alcanzar. La forma de determinarlas forma parte del algoritmo. Estas se definen de acuerdo con las metas de negocio, el perfil del navegador y su historia o comportamiento pasado. De esta manera, es posible, que una página sea página objetivo en una visita de un usuario al sitio web y no sea parte del conjunto de páginas objetivo en una posterior visita del usuario al mismo sitio web.

Enlace: Un enlace es un arco con origen en la página a_i y destino en la página a_j . Los enlaces se representan por medio del par (a_i, a_j) .

Valor de un enlace: La principal acción del usuario es seleccionar un enlace para obtener la siguiente página (o terminar la sesión). Esta acción toma diferentes valores dependiendo de la distancia o cercanía a la página o conjunto de páginas objetivo.

El valor del enlace (a_i, a_j) se representa por medio de un número real v_{ij} ($v_{ij} \in \mathbb{R}$ for $0 \leq i, j \leq n$):

- Si $v_{ij} > 0$, consideramos que el navegante, yendo del nodo i al nodo j , está más cerca de las páginas objetivo. (Si $v_{ij} > 0, v_{il} > 0, v_{ij} > v_{il}$ entonces, se considera que es mejor ir de la página α_i a la α_j que ir de la página α_i a la α_l)
- Si $v_{ij} < 0$ se considera que el navegante, que va de la página α_i a la α_j , se está alejando de las páginas objetivo. (Si $v_{ij} < 0, v_{il} < 0, v_{ij} < v_{il}$ entonces es peor ir de la página α_i a la página α_j que ir de la α_i a la α_l)
- Si $v_{ij} = 0$ consideramos que el enlace no representa ni una ventaja ni una desventaja en la búsqueda del objetivo.

Sesión: es una secuencia de páginas visitadas por un usuario. El archivo de registro de accesos al sitio web contiene información relacionada con la identificación del usuario (dirección IP), URL de la página solicitada y fecha y tiempo de la solicitud. Con esta información se puede reconstruir la sesión representada como un vector de páginas recorridas: $S[1], S[2], \dots, S[n]$.

Las sesiones se denotan por S siendo $|S|$ su longitud (número de páginas visitadas). Las sesiones se representan como vectores de manera que $S[i]$ representa la i -ésima página visitada $S[i] \in \mathbf{W}$ $1 \leq i \leq n$, con, $|S| = n$.

Es importante destacar que las páginas del sitio web visitadas durante una sesión se pueden repetir. Por ejemplo, si la primera y sexta páginas visitadas son la página 3, $S[1]=S[6]=a_3$. Sin embargo, las páginas contenidas en el sitio web a_1, \dots, a_m no se repiten dado que conforman un conjunto.

Secuencia inicial de longitud k : (k páginas iniciales $S[1], S[2], \dots, S[k]$): las primeras k páginas recorridas durante una sesión representan una secuencia de los $k-1$ enlaces iniciales de la sesión.

Valor de una secuencia inicial de longitud k : $S[1], S[2], \dots, S[k]$: este valor se calcula como la suma de cada uno de los valores de las páginas $S[k]$ a las cuales llega el usuario recorriendo los enlaces $(S[1], S[2]), (S[2], S[3]), \dots, (S[k-1], S[k])$ y se denota por $AV(k)$ (Added Value of a k -length Initial Sequence).

$$AV(k) = v_{S[1], S[2]} + v_{S[2], S[3]} + \dots + v_{S[k-1], S[k]} \quad \text{Ecuación 2}$$

El valor acumulado de una secuencia inicial se puede definir como:

$$AV(k) = AV(k-1) + v_{S[k-1], S[k]}$$

Valor de Sesión: Se calcula como la suma de los valores de los enlaces recorridos durante una sesión completa (páginas visitadas) y se denota por $VA(n)$, donde

$$AV(n) = v_{S[1], S[2]} + v_{S[2], S[3]} + \dots + v_{S[n-1], S[n]} \quad \text{Ecuación 3}$$

Valor promedio de una sesión: éste representa el valor promedio de cada enlace recorrido durante una sesión. Denotado por $AAV(n)$ - Average session value -, se define como el valor total de la sesión dividido por el número total de enlaces recorridos. El número de enlaces recorridos es $n-1$, al final de una sesión en la cual se han recorrido n páginas.

$$AAV(n) = AV(n) / (n-1)$$

Interpretación del Valor promedio acumulado de una secuencia inicial de longitud k ($AAV(k)$): Este valor ofrece al administrador del sitio web el valor promedio generado para cada uno de los enlaces recorridos hasta alcanzar la página k -ésima. Si ejecutáramos el algoritmo durante una sesión (en tiempo real) obtendríamos una medida útil que es independiente del número de enlaces recorridos. Por ejemplo, si tenemos páginas web adaptativas (tienen en cuenta diferentes parámetros) podríamos ofertar productos y/o páginas más atractivas a aquellos usuarios con un bajo Valor Acumulado en una secuencia inicial de longitud k . De esta manera, se podría incrementar el valor promedio acumulado de cada usuario en una sesión.

- Si el número de páginas recorridas (k) se incrementa, el Valor Acumulado de una secuencia inicial promedio en las k páginas iniciales ($AAV(k)$), puede:
 - **Incrementar**, lo que significa que la sesión se acerca la meta.

- **Decrementar** : cuando la sesión se aleja de la meta
- **Permanecer constante** : cuando la sesión ni se acerca ni aleja de la meta

3. Algoritmo para el cálculo del valor de una sesión

El algoritmo tiene por objetivo conocer lo cercano que está el comportamiento de un usuario del sitio web de los objetivos de la organización. Medimos la distancia de los objetivos utilizando el valor de los enlaces recorridos. El algoritmo de cálculo de evolución del valor de una sesión se basa en el recorrido de un grafo.

La entrada es una matriz de valores $V[m,m]$ que contiene el valor de los enlaces en un sitio web, que se determinan en base a los procesos de negocios de la organización y los objetivos del sitio web. El análisis de los procesos de negocios brinda un marco conceptual para determinar el valor de los enlaces, en función de cuanto acercan (o alejan) al usuario de las páginas establecidas como objetivo del sitio web.

Las matrices de valoraciones son, en consecuencia, determinadas por los directivos de negocios de la organización. Es de destacar que pueden ser calculadas para cada perfil de usuario y por tanto hacen posible adaptar los objetivos empresariales de acuerdo al comportamiento de los usuarios.

Como consecuencia, las matrices de valoraciones V son adaptables en dos aspectos:

- Los objetivos empresariales no son fijos ni únicos, pueden modificarse o ser considerados desde diferentes puntos de vista. Por ejemplo marketing, ventas, auditoria y relaciones públicas pueden analizar una misma sesión desde sus puntos de vista. Para reflejar diferentes puntos de vista el algoritmo deberá ejecutarse con diferentes matrices de valoraciones como input.

- Por otro lado, la entrada del algoritmo puede incorporar (además de la matriz de valoraciones M) la identificación del usuario. La matriz de valores de los enlaces puede adaptarse en función del perfil del usuario, definido por ejemplo en un proceso previo de segmentación de los usuarios.

Las salidas del algoritmo son la evolución del valor acumulado y del valor acumulado promedio durante la sesión.

Pseudocódigo del algoritmo

Input: Value links matrix $V[m,m]$

Inicialización

```
AV=0 //Added Value=0
AAV=0 //Average added value=0
k=1 //number of nodes=1
read S[k] //read the first traversed page in the Web site S[1]
```

Pseudocode:

```
While new pages are traversed
    k = k + 1 //compute the traversed page sequential number
    read S[k] // read the next traversed page
/* the selected link is (S[k-1],S[k]) 1≤S[k-1]≤m-1 1≤S[k] ≤m
2≤k≤n */
AV = AV + V(S[k-1],S[k]) // Add link traversed value to accumulated value
AAV = AV / k-1
// Compute average link traversed value
Plot values
// needs the previous value to be stored
```

Output: Final Accumulated Value and Final Average Accumulated Value

/* or Accumulated Value and Average Accumulated Value evolution if all values are stored. */

4. Valor de una Sesión

La principal ventaja del algoritmo propuesto es el cálculo de la evolución del valor de una sesión., que constituye un elemento relevante en la adopción de decisiones de diseño relativas a sitios y páginas web.

Una empresa puede beneficiarse con estos resultados al detectar la necesidad de incorporar nuevas páginas, realizar ofertas online o efectuar ventas cruzadas. Frecuentemente, los ejecutivos no comprenden cómo usar la tecnología y qué tipo de análisis efectúa [3]. Nuestra propuesta sugiere adaptar la tecnología a las métricas de los ejecutivos [15]. Su principal ventaja es que no requiere de grandes esfuerzos para entenderla y por tanto el esfuerzo requerido para utilizarla se minimiza.

A continuación presentamos ejemplos de análisis del valor de una sesión seguidos de un ejemplo que ilustra el comportamiento del algoritmo.

4.1. Análisis de la evolución del valor de una sesión

Para analizar la evolución del valor de una sesión se presenta una gráfica en la cual las abcisas representan la cantidad de enlaces recorridos y las ordenadas el valor acumulado hasta el último enlace recorrido.

En la Figura 1 se muestra la evolución del valor de una sesión, que disminuye al comienzo y que luego aumenta. Durante esta sesión el usuario se fue alejando del objetivo y posteriormente volvió a él.

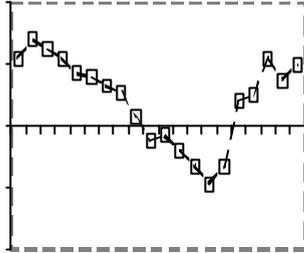


Figura 1 Evolución del valor acumulado de una sesión

Las Figuras 2-3 describen tres sesiones que tienen similar evolución de valor. De esta manera es posible encontrar dos patrones en las seis sesiones .

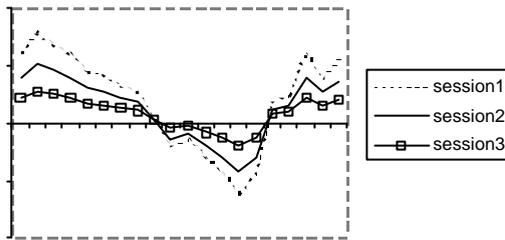


Figura 2- Evolución del valor acumulado (sesiones 1 a 3).

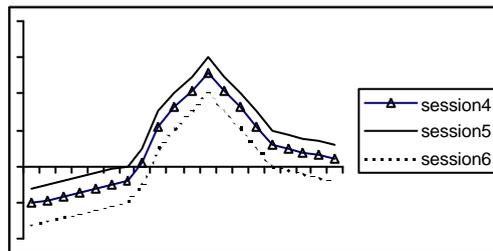


Figura 3-Evolución del valor acumulado (sesiones 4 a 6)

A partir de las sesiones 1 a 6 que aparecen en la Figura 4 se pueden obtener dos patrones interesantes.

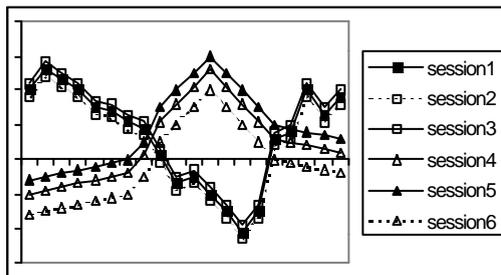


Figura 4 - Evolución del valor acumulado (sesiones 1 a 6)

Combinando el análisis de la evolución del valor con las características de las páginas del sitio web es posible identificar las páginas que llevan a secuencias ascendentes o descendentes con respecto al valor de la sesión. En particular, estas páginas serían las previas o iniciales de ciclos de valores descendentes. Esto puede permitir detectar aspectos a modificar en las mismas.

4.2. Cálculo del valor de una sesión

En esta sección se presenta un ejemplo de cálculo del valor de una sesión. El ejemplo permite observar la baja complejidad del algoritmo propuesto. En lugar de constituir una limitación, su simplicidad puede ser considerada un elemento positivo, porque facilita la comprensión de sus resultados por parte de los administradores de un sitio web.

La entrada del algoritmo es la siguiente matriz de valores de los enlaces $V[4,4]$

	Origen (α_i)			
Destino (α_j)	α_1	α_2	α_3	α_4
α_1	3	2	3	6
α_2	4	1	2	2
α_3	-5	-1	-1	-1
α_4	-6	-2	-3	-1

En la Figura 5 se muestra el grafo con la asignación de valores a los arcos asociada a la matriz anterior. Es fácil observar que en el grafo existen dos nodos objetivo (α_1 y α_2), porque los arcos de entrada a estos nodos son positivos y los arcos que salen hacia el resto de los nodos (α_3 y α_4) toman valores negativos.

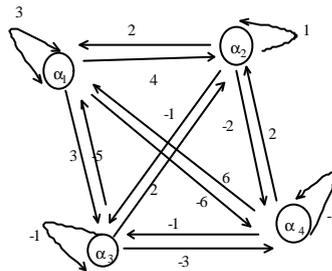


Figura 5- Grafo con los valores de los enlaces

Los enlaces que coinciden en origen y destino, que se observan en la diagonal de la matriz de valoraciones y en el grafo correspondiente, representan la acción de *reload* (volver a cargar) una página.

5. Determinación de Matrices de Valor de los enlaces

La metodología utilizada para determinar las matrices de valores puede subdividirse en dos fases:

- El primer paso es definir un algoritmo para determinar preferencias explícitas. Se basa en una etapa de identificación de las mismas y en el análisis de las respuestas a los formularios diseñados a tales efectos.
- El segundo paso consiste en la determinación de matrices de preferencias implícitas, utilizando las técnicas de estimación de data mining. La estimación utiliza las matrices de preferencias explícitas obtenidas en el primer paso y en el análisis de las asignaciones de recursos efectuadas y las decisiones adoptadas.

La motivación para la incorporación del segundo paso es la siguiente. Si por ejemplo, en el paso 1 se clasifica una página como muy importante, pero luego se detecta que no se destinan recursos humanos ni financieros a ella, implícitamente no es tan importante como se dice. En función de la utilización de recursos, se determina un modelo para estimar su valor implícito. Si por el contrario, se cataloga como no importante a una página a la que se destinan cuantiosos recursos, inferimos que es más importante de lo que se planteó inicialmente.

6. Resultados Experimentales

El análisis y las pruebas se hicieron con datos de 2400 sesiones de usuario provenientes de un sitio web de comercio electrónico. Los datos fueron procesados teniendo en cuenta 4 matrices diferentes. Se descartaron las sesiones que recorrían 10 páginas o menos, dado que la propuesta no resulta interesante para analizar sesiones cortas.

Los valores de las cuatro matrices utilizadas premian los siguientes aspectos:

1. área de noticias ,
2. página principal
3. registro de usuarios
4. realización de compras .

En la figura siguiente se muestra que al analizar los valores de los acumulados por sesión con las cuatro matrices se observan patrones claramente definidos. Se utiliza una matriz de diagramas de dispersión, que muestra todas las combinaciones posibles de los valores acumulados por sesión obtenidos con las cuatro matrices consideradas. La primera fila y la primera columna presentan cálculos obtenidos con la matriz 1, la segunda fila y la segunda columna los obtenidos con la matriz 2, y así sucesivamente.

En el cruce de la primera fila y la segunda columna se presenta el gráfico que vincula valores de sesión obtenidos con las matrices 1 y 2. En el cruce de la segunda fila y la primera columna se presenta (con los ejes invertidos) el análisis de las mismas matrices. No se cruzan los valores obtenidos con una matriz, consigo misma, dado que siempre el gráfico estará formado por puntos ubicados en la diagonal principal. Al analizar el gráfico, se observa por ejemplo como los valores obtenidos con las matrices 1 y 2 muestran son totalmente opuestas.

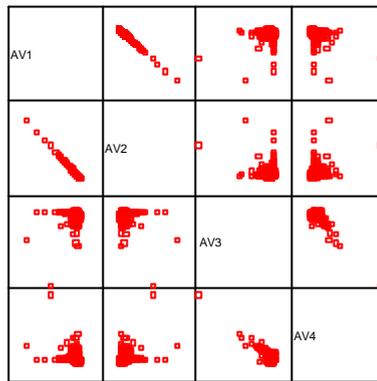


Figura 6 - Valores acumulados con las 4 matrices consideradas.

En la Figura 7 se muestran los valores promedio obtenidos con las cuatro matrices consideradas. Se observa el mismo patrón que para los valores acumulados.

Con el objetivo de resumir los valores promedio de cada enlace recorrido en una sesión obtenidos con las cuatro matrices, calculamos media y desviación estándar, que se presentan en la Tabla 1. El valor positivo +4,8048 muestra que las sesiones son favorables al ser analizadas con el criterio subyacente a la matriz 2. Sin embargo, el valor promedio -3.3382 muestra una evaluación desfavorable si se analizan las sesiones con el criterio que brinda la matriz 1.

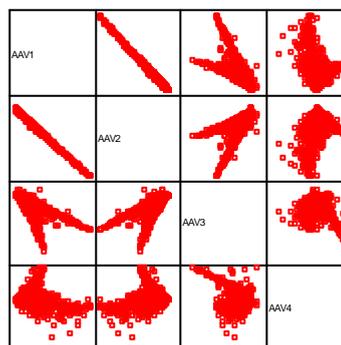


Figura 7 - Valores promedio con las 4 matrices consideradas

Al observar los valores de resumen para el valor promedio al final de la sesión, se observa que los caminos, en promedio, se adecuan a los objetivos reflejados en la matriz 2 y no se adecuan a los planteados en los objetivos reflejados en la matriz 1

Matriz	Media	Desviación Estandar
1	-3,3382	2,1435

2	4,8048	2,8445
3	-,8720	1,2446
4	,1187	,3767

Tabla 1 -Media y desviación estándar del valor acumulado promedio

Por último, los patrones que se obtienen con las diferentes matrices se presentan en la Figura 8

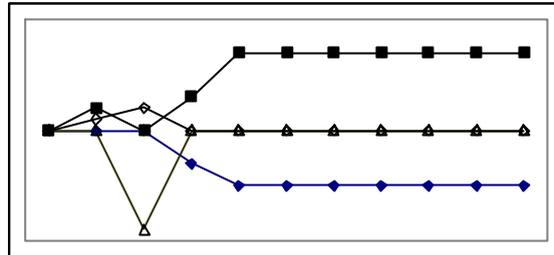


Figura 8 Valores obtenidos con datos experimentales.

7. Conclusiones y trabajo futuro

En este artículo se ha presentado un enfoque para calcular el valor de una sesión de usuario en un sitio web teniendo en cuenta tanto su comportamiento como las metas establecidas por el administrador del sitio web.

El enfoque propuesto ha sido probado con datos reales y demuestra ser útil para seguir el comportamiento de los usuarios mientras navegan en un sitio. Por otra parte, del análisis a posteriori de las sesiones, es posible identificar sesiones similares incluso cuando no se accede a las mismas páginas. Asimismo, la posibilidad de definir distintos tipos de matrices, de acuerdo a distintos criterios posibilita el poder realizar análisis bajo diferentes puntos de vista.

El mayor problema del algoritmo radica en el hecho de que se requieren las matrices de valoración de entrada y de momento éstas se calculan manualmente. Estamos trabajando en la actualidad en un prototipo para el cálculo de estas matrices aplicando técnicas de data mining, del cual se presentó un esbozo en la sección 5

Otro problema del algoritmo propuesto es que no tiene en cuenta los tiempos de permanencia del usuario web en una página. En la siguiente versión del algoritmo está prevista una modificación para tener en cuenta estos tiempos.

Agradecimientos

Esta investigación ha sido parcialmente financiada por el proyecto financiado por la Universidad Politécnica de Madrid: WEB-RT: Un sistemas de detección de patrones en la web en tiempo real

Bibliografía

- [1] B. Mobasher, N. Jain, E. Han, and J. Srivastava. (1997) "Web mining: Pattern discovery from WWW transaction". In *Int Conference on Tools with Artificial Intelligence*, pages 558-567, New port.
- [2] J. Han, M. Kamber. *Data Mining: Concepts and Techniques*. Academic Press USA 2001
- [3] G. Piatetsky-Shapiro "Interview with Jesus Mena, CEO of WebMiner, author of *Data Mining your Website*" Date: June 24, 2001 <http://www.kdnuggets.com/news/2001/n13/13i.html>
- [4] J. Borges and M. Levene. "Mining navigation patterns with hypertext probabilistic grammars" *Research Note RN/99/08*, Department of Computer Science - University College London, 1999.
- [5] J. Borges and M. Levene. "Data mining of user navigation patterns". *Web Usage Mining, in Lecture Notes in Artificial Intelligence (LNAI 1836)* B. Masand and M. Spliliopoulou, editors., Springer-Verlag, Berlin, 2000.
- [6] M. Perkwitz and O. Etzioni "Adaptive Web Sites: Automatically Synthesizing Web Pages". In *Proceedings of AAAI98*.
- [7] M. Perkwitz and O. Etzioni. "Adaptive web sites: Conceptual cluster mining". In *Sixteenth International Joint Conference on Artificial Intelligence*, Stockholm, Sweden, 1999.
- [8] M. Perkwitz and O. Etzioni. "Towards adaptive Web sites: Conceptual framework and case study". In *Artificial Intelligences* 118, 2000.
- [9] M. Spliliopoulou, L. Faulstich, and K. Wilkler. "A data miner analyzing the navigational behaviour of web users". In *Proc. Of the Workshop on Machine Learning in User Modelling of the ACAI99*, Greece.

- [10] M. Spiliopoulou, C. Pohle, and L. Faulstich. "Improving the effectiveness of a web site with web usage mining". In *Proceedings WEBKDD99*.
- [11] R. Cooley, B. Mobasher, and J. Srivastava. "Data preparation for mining world wide web browsing patterns". *Knowledge and Information Systems*, 1(1), February 1999.
- [12] E. Menasalvas, S. Millán, J.M. Peña, M. Hadjimichael, O. Marbán " Subsessions: A granular approach to click path analysis" In *Proceedings WCCI'2002*
- [13] C. Shalabi, F. Banaei-Kashaani, J. Faruque and A. Faisal. (2001) "Feature matrices: A model for e-Ecient and anonymous web usage mining". In *Proceedings of EC-Web 2001*, Germany, September
- [14] E. Menasalvas, O. Marbán , S. Millán , and J. M. Peña "Intelligent Web Mining" in *Intelligent Exploration of the Web series Studies in Fuzziness and Soft Computing*., Springer-Verlag 2002 P.S.Szczepaniak, J. Segovia, J. Kacprzyk, L.A. Zadeh – editors
- [15]. E. Menasalvas , E. Hochztain . "Sessions value as measure of web site goal achievement" .. In *Proceedings of SNPD '02 - Madrid (Spain), June, 2002*