

# Evaluación en entornos de recomendación mediante técnicas de aprendizaje automático

Alejandro Bellogín Kouki  
Universidad Autónoma de Madrid  
alejandro.bellogin@uam.es

19 de mayo de 2008

## 1. Introducción

La evaluación de un entorno adaptativo es una tarea compleja y, normalmente, costosa, debido sobre todo al gran número de factores que influyen en la adaptación, además de la componente subjetiva que contienen la mayoría de estos sistemas. Esto se hace más notorio en el caso de los sistemas de recomendación.

En este trabajo aplicaremos ideas utilizadas con anterioridad en el campo de las aplicaciones adaptativas (más concretamente, en el área de la educación) para evaluar los resultados de un sistema de recomendación y descubrir cuáles son los parámetros del sistema y/o del usuario que facilitan las labores de un algoritmo de recomendación.

Se presentará un sistema que permite personalizar de acuerdo al contenido de un perfil, crear un contexto a partir de clicks del usuario y de las últimas noticias vistas y en el que se puede activar o desactivar la expansión semántica de las preferencias del usuario. Además, se comprobará empíricamente la certeza de dos hipótesis:

1. La expansión mejora la personalización.
2. Los mejores resultados se encuentran al activar personalización y contexto.

La evaluación utiliza datos reales extraídos a partir de una interacción controlada por parte de 16 usuarios con el sistema.

## 2. Trabajos similares

En este trabajo se utilizan las técnicas que se usan en [8] pero aplicado a otro tipo de entorno adaptativo: un sistema de recomendación. Este sistema se explicará con más detalle en la sección 3. El sistema se someterá a varios niveles de evaluación, y utilizaremos las herramientas que el aprendizaje automático pone a nuestra disposición (sección 4) para extraer patrones interesantes de esos datos. De esta manera, se pueden conseguir los siguientes objetivos:

- Confirmar las hipótesis de partida
- Descubrir parámetros que influyen en la precisión de la recomendación

No se conoce de ningún autor que haya realizado una evaluación similar (aprendizaje automático aplicado a un sistema de recomendación) a la que se propone aquí.

Hay que comentar, que los algoritmos de aprendizaje se aplicarán sobre los logs que el sistema recogió después de una evaluación de dos días con 16 usuarios. La evaluación por parte del usuario del sistema es muy parecida a la realizada en [1], en parte porque los dos sistemas de recomendación tienen el mismo dominio: noticias. En [7] se puede ver una evaluación de un sistema adaptativo que utiliza contexto.

Los resultados encontrados al aplicar las distintas técnicas automáticas se pueden encontrar en la sección 5. En las últimas secciones analizaremos de qué manera podría continuarse este trabajo y qué conclusiones se sacan del mismo.

### 3. Arquitectura del sistema

El sistema adaptativo que vamos a estudiar permite recomendar noticias, utilizando distintos algoritmos, más concretamente:

- Orientado al individuo
- Orientado a grupo
  - Colaborativo: basado en usuario o en ítem
  - Híbrido
- Sensible al contexto

Detallaremos a continuación algunos aspectos relevantes del sistema, para más detalles se pueden consultar [2, 3] para saber más cosas acerca del sistema y [2, 4] sobre los modelos de recomendación.

Existe una parte de la interfaz que permite al usuario controlar los algoritmos a utilizar y el peso que las preferencias personales (a largo y corto plazo) tendrán sobre las recomendaciones.

Este sistema tiene control sobre el vocabulario que se utiliza para describir tanto el contenido de las noticias como las preferencias del usuario. Para ello, utiliza una ontología, aprovechando las relaciones entre conceptos en el proceso de recomendación y teniendo en cuenta que el perfil del usuario y las noticias están representados por medio de estos conceptos. Las noticias se introducen en el sistema automáticamente y periódicamente mediante *agregadores* RSS. El sistema almacena su título, resumen y categoría a la que pertenece para anotarlas automáticamente después y mostrarlas en una de las ocho categorías disponibles en el sistema.

El proceso mediante el cual un usuario describe sus preferencias a través de conceptos de la ontología es manual. En la figura 2 se puede ver el editor de perfiles que permite esta funcionalidad. En él, el usuario puede buscar conceptos y añadirlos a su perfil, asignando un peso a dicho concepto (si el peso es negativo, se entiende que el usuario posee un cierto grado de rechazo a ese concepto, lo cual se traducirá en que las noticias que contienen ese concepto aparecerán las últimas en el sistema). Si está muy interesado en un concepto o prefiere hacer una distinción más clara o crear un perfil más definido, también es posible añadir al perfil *instancias* de un determinado concepto, es decir, ejemplos del mundo real que se pueden considerar subtipos del concepto. Además, a través de los ratings, etiquetas y otras acciones del usuario se infieren preferencias implícitas del usuario, que junto con el historial de clicks define sus preferencias a corto plazo, o *contexto*.

Por otra parte, la asignación de conceptos a las noticias se realiza mediante técnicas de Procesamiento de Lenguaje Natural, enriqueciendo una ontología con instancias extraídas a partir de categorías de la Wikipedia<sup>1</sup> [5]. En nuestro caso la ontología utilizada es la unión de 17 ontologías, adaptaciones de la ontología IPTC<sup>2</sup>, la cual contiene conceptos de dominios diversos, tales como arte, ciencia, cultura, educación, política o religión. En la figura 1 se puede ver una parte de la ontología.

---

<sup>1</sup><http://www.wikipedia.org/>

<sup>2</sup>Disponible en <http://nets.ii.uam.es/neptuno/iptc/>

Desde el punto de vista técnico, el sistema se basa en una arquitectura cliente / servidor y utiliza la tecnología AJAX<sup>3</sup> para presentar de manera dinámica la interfaz gráfica adaptativa.

Por último, de cara a la evaluación, se añadió a la interfaz la posibilidad de expandir o no las preferencias del perfil. Esto afecta a la hora de calcular el valor de personalización asociado a cada noticia, ya que un perfil expandido es más probable que coincida con las anotaciones de una noticia que uno sin expandir.

A continuación, describiremos la manera en que se puede evaluar un sistema con estas características, y cómo se puede aplicar el aprendizaje automático para extraer información acerca del comportamiento del sistema.

### 3.1. Evaluación del sistema

Dado el gran número de algoritmos involucrados en el sistema, su evaluación resulta muy complicada. El método más común de evaluar un sistema de personalización es comparar los resultados obtenidos con la personalización activada frente a cuando está desactivada y utilizando cuestionarios [1]. Por un lado, estos métodos no dicen cuánto de bueno es el algoritmo de personalización; por otro, no son útiles en nuestro caso, ya que tendríamos que realizar  $2^n$  pruebas con  $n$  algoritmos o métodos distintos, sin ninguna información global que dé coherencia a los resultados.

Por ello, la manera de evaluación que hemos seguido ha sido la de probar varios métodos de una sola vez (en la misma tarea), de manera que se pueda ver qué situaciones son las más favorables para con el usuario, al comparar con las tareas en que los algoritmos están desactivados.

Actuando de esta forma, tenemos los siguientes casos:

1. No personalización
2. Sólo contexto (con expansión semántica)
3. Personalización con expansión semántica y contexto
4. Personalización sin expansión semántica
5. Personalización con expansión semántica, contexto, filtrado colaborativo
6. Personalización sin expansión semántica, contexto, filtrado colaborativo
7. Personalización con expansión semántica y filtrado colaborativo
8. Personalización sin expansión semántica y filtrado colaborativo
9. Personalización con expansión semántica y contexto orientado a grupo

Se pueden plantear las siguientes hipótesis:

- La expansión mejora la personalización ya que aumenta las probabilidades de encontrar una noticia relacionada con las preferencias
- La utilización de contexto junto con la personalización mejora la personalización al tener en cuenta tanto las preferencias a largo plazo (perfil) como a corto (contexto)

El objetivo principal de este trabajo es utilizar técnicas de aprendizaje automático para corroborar que estas hipótesis son ciertas sobre datos extraídos de evaluaciones con 16 usuarios reales.

En dicha evaluación se realizaron únicamente pruebas de los cuatro primeros casos, debido a que era una evaluación preliminar del sistema y, además, evaluar grupos y filtrado colaborativo requeriría que todos

---

<sup>3</sup>Asynchronous JavaScript And XML



Figura 1: Parte de la estructura de la ontología IPTC

My preferences

Concept	Weight	Public for		
		All people	My contacts	Only me
football	<input type="text" value="100"/>	<input checked="" type="radio"/>	<input type="radio"/>	<input type="radio"/>
soccer	<input type="text" value="100"/>	<input checked="" type="radio"/>	<input type="radio"/>	<input type="radio"/>
cinema industry	<input type="text" value="100"/>	<input checked="" type="radio"/>	<input type="radio"/>	<input type="radio"/>

Insert a defined concept or select it from the categories

cinema industry

Vocabulary about economy, business and finance

- economy, business and finance**
- transport
- metal and mineral
- consumer goods
- construction and property
- energy
- metal goods
- computing and information technology
- company information
- media
- book industrv

- a list
- a-list
- abraham van helsing
- adam buxton
- adam west
- aerial shot
- agent smith
- akira kobayashi
- al franken
- alain delon
- alan price
- alastair duncan

This system has been developed by Iván Cantador and Alejandro Bellogín (Networked Semantics Team, <http://nets.ii.uam.es>)

Figura 2: Editor de perfiles de News@Hand

los usuarios hubieran hecho una evaluación completa (primera fase). En estos momentos, se está llevando a cabo la segunda fase que comprende evaluación del resto de los algoritmos de recomendación presentes en el sistema (colaborativo, híbrido y grupos).

Hay que notar que en estas situaciones no es necesaria la existencia de una consulta, aunque para la definición de una tarea concreta orientada a evaluación suele ser útil tener una consulta o tarea predefinida, con la cual el usuario pueda crear una situación real y evaluar el sistema en consecuencia. En la evaluación que se llevó a cabo, la consulta servía para centrar las noticias alrededor de un tema, pero además, los usuarios tenían un perfil predeterminado y un objetivo a corto plazo. Por ejemplo, una tarea real del sistema consistía en marcar como relevantes noticias que mencionaran el tema de la *clonación* o los *animales* (objetivo), para ello las noticias se habían filtrado (se había realizado una búsqueda) por el término *comida*, el usuario tenía preferencias en su perfil acerca de *salud*, *gente joven* y evaluaba 15 noticias atendiendo al siguiente criterio: en una escala del 1 al 5 cada noticia la puntuaba

- Como 1 si no tenía que ver ni con el objetivo ni con el perfil
- Como 3 si sólo tenía que ver con el objetivo
- Como 4 si tenía relación con el perfil
- Como 5 si hablaba de conceptos relacionados tanto con el objetivo como con el perfil

En particular, los 16 usuarios llevaron a cabo 4 tareas, dos de ellas se dividían en dos fases y otras dos, en tres fases.

A partir de todos estos datos, se aplican técnicas de aprendizaje automático para averiguar cuáles son las características que más han influido para la obtención de noticias relevantes y no relevantes. Para ello, las características o atributos que se van a tener en cuenta para cada juicio de relevancia (cada noticia con una puntuación determinada) son:

- ¿La personalización estaba activada?
- ¿El contexto estaba activado?
- ¿El perfil se generó de manera automática?
- ¿La expansión estaba activada?
- Tamaño de contexto
- Tamaño de perfil
- Número de página en que apareció la noticia
- Número de noticias abiertas por el usuario en esa sesión

En la figura 3 se puede ver una imagen de la cabecera del fichero en formato ARFF usado por los algoritmos de aprendizaje automático descritos en la sección 4.

Además la evaluación se va a plantear desde tres puntos de vista distintos:

- General. Las noticias relevantes (clase positiva en el algoritmo de aprendizaje automático) serán aquellas en que la puntuación sea mayor o igual que 3, es decir, cualquiera que tuviera que ver con el objetivo o con el perfil.
- Orientada a perfil. En este caso las noticias relevantes sólo serán aquellas que tuvieran relación con el perfil, es decir, las puntuadas como 4 o 5.

```

@attribute personalisation      {YES, NO}
@attribute context              {YES, NO}
@attribute manualPreferences    {YES, NO}
@attribute expansion            {YES, NO}
@attribute contextSize         {None, Small, Medium, Large}
@attribute profileSize         {None, Small, Medium, Large}
@attribute topicSection        {Headlines, World, Business, Technology, Science, Health,
@attribute rankingPage         {1, 2, 3, OTHER}
@attribute contextPhase        {-1, 0, 1, 2}
@attribute class                {LIKE, DONTLIKE}

@data

```

Figura 3: Cabecera de los ficheros ARFF usados para aprendizaje automático

- Orientada a objetivo. Las noticias relevantes son aquellas con puntuaciones 3 o 5.

Distinguiendo estos tres casos, será posible encontrar características que influyen de manera general en noticias relevantes y no relevantes, pero además, se podrá encontrar, por ejemplo, características muy correlacionadas con el perfil o con el objetivo.

A continuación se explicarán las técnicas del área de aprendizaje automático que se han utilizado y en la siguiente sección se mostrarán los resultados encontrados a partir de la evaluación y análisis recién explicados.

## 4. Técnicas empleadas

En esta sección explicaremos brevemente las técnicas que se mencionan en la sección 5, para más detalles se pueden consultar [9] y [6].

El objetivo de este trabajo es comprobar empíricamente que se cumplen determinadas hipótesis y descubrir dependencias entre atributos del sistema y los resultados de los algoritmos de recomendación. Por ello, los algoritmos involucrados en esta evaluación no se utilizarán nunca con datos nuevos, sino que nos servirán para clasificar datos que ya se conocen y de los cuales se quieren extraer relaciones.

La primera técnica que se utilizará es la de los árboles de decisión. Estos árboles dan una aproximación del tipo *divide y vencerás* a los problemas de aprendizaje, en la manera en la que en cada nodo, se compara uno de los atributos y se ramifica según el valor de dicho atributo. Si el árbol no es muy grande, se obtiene mucha información con su representación visual, lo cual es una ventaja frente a otros algoritmos de clasificación, en donde hay que representar curvas de error o similares para hacerse una idea del funcionamiento del algoritmo (léase redes neuronales o técnicas estadísticas).

Las reglas de clasificación son una buena alternativa a los árboles de decisión cuando se repiten las comprobaciones a lo largo del árbol, ya que las reglas no entienden de orden y permiten ser más compactas. Mantiene la ventaja de ser comprensibles por los humanos y, en el caso de clases binarias, permitir la asunción de mundo cerrado simplifica aún más el número de reglas.

Una última pero muy interesante aplicación de las técnicas de aprendizaje automático a los datos analizados es la de la selección de atributos, de esta manera se pueden descubrir, dado un conjunto de datos, cuáles son los atributos que mayor información almacenan, obteniendo un orden de importancia entre ellos o subconjuntos que estén muy correlacionados con la clase objetivo. En este trabajo, se ha utilizado como medida de información la ganancia de información, y como método de búsqueda en el espacio de atributos un algoritmo *greedy* (voraz) con descenso de gradiente.

En el caso estudiado, no se han utilizado las reglas de asociación, a diferencia de [8], ya que sólo interesaba predecir las clases, no obstante, hay que mencionar que son una herramienta de clasificación muy

potente, similar a las reglas de clasificación pero que pueden predecir cualquier atributo o combinación de atributos.

La aplicación de estas técnicas se ha realizado a través de la herramienta Weka, colección de algoritmos de aprendizaje automático de código abierto, que permite tanto aplicarlos directamente mediante una interfaz gráfica como ser llamados desde código Java. Permite preprocesamiento, clasificación, regresión, clustering, reglas de asociación y visualización de resultados. En este trabajo se utilizó la versión 3.5.7 disponible en [www.cs.waikato.ac.nz/ml/weka](http://www.cs.waikato.ac.nz/ml/weka). Los parámetros de los distintos algoritmos empleados en este trabajo se encuentran en la tabla 1.

Algoritmo	Parámetros	Funcionalidad
GainRatioAttributeEval	Ranker -T -1.7976931348623157E308 -N -1	Ordenación de atributos
CfsSubsetEval	BestFirst -D 1 -N 5	Selección de atributos
J48	-C 0.3 -M 5	Clasificación por árbol de decisión
PART	-M 2 -C 0.35 -Q 1	Generación de reglas de clasificación

Cuadro 1: Parámetros de los algoritmos utilizados en la herramienta Weka

## 5. Resultados

De los datos extraídos de la evaluación (sección 3.1) se extrajo la información relevante explicada con anterioridad (figura 3). En este momento se tienen tres ficheros, uno por cada tipo de evaluación que se quiere estudiar: general, orientada a perfil y orientada a objetivo<sup>4</sup>.

Al aplicar distintas técnicas de aprendizaje automático a cada fichero se obtiene información distinta. Por ejemplo, se pueden ver los árboles de decisión generados por cada tipo de evaluación en las figuras 4, 5 y 6. A primera vista, se puede distinguir la complejidad que involucra cada tipo de evaluación.

Como se ha podido apreciar, los árboles de decisión están muy influidos por la sección en donde aparece la noticia. Esto es un indicador de que la evaluación no ha sido todo lo uniforme que debería, ya que hay secciones donde es más fácil encontrar noticias relevantes que en otras (hay que recordar que cada fase de la evaluación tenía en cuenta el perfil, una consulta para restringir las noticias a evaluar y un objetivo, además de parámetros relacionados con los algoritmos de recomendación).

No obstante, se pueden extraer algunas conclusiones de los árboles presentados:

- En la evaluación general se ven dos secciones que se clasifican automáticamente sin necesidad de estudiar más atributos. En el resto de secciones, se pueden descubrir ciertas constantes en los datos: las preferencias manuales no favorecen el hecho de encontrar noticias relevantes, en la primera página se encuentran noticias relevantes siempre y cuando esté activa la personalización. Sin embargo, mientras que en la sección de Economía la activación del contexto ayuda a encontrar noticias relevantes, en la sección de Entretenimiento ocurre justo lo contrario. Con respecto a la expansión, sólo aparece en una de las ramas del árbol, indicando que sólo es útil si se utiliza junto con la personalización.
- En la evaluación orientada a perfil sólo hay una sección (Internacional) que no necesita más atributos para ser clasificada. Sobre el resto de secciones, en Portada, sólo son relevantes las noticias que aparecen en la primera página si está activada la personalización, en Economía es importante que el contexto esté activado, mientras que si no lo está, obtienen mejores resultados las preferencias no

<sup>4</sup>En realidad la información contenida en estos ficheros es igual excepto la clase de cada instancia, que cambia según se explicó en la sección 3.1.

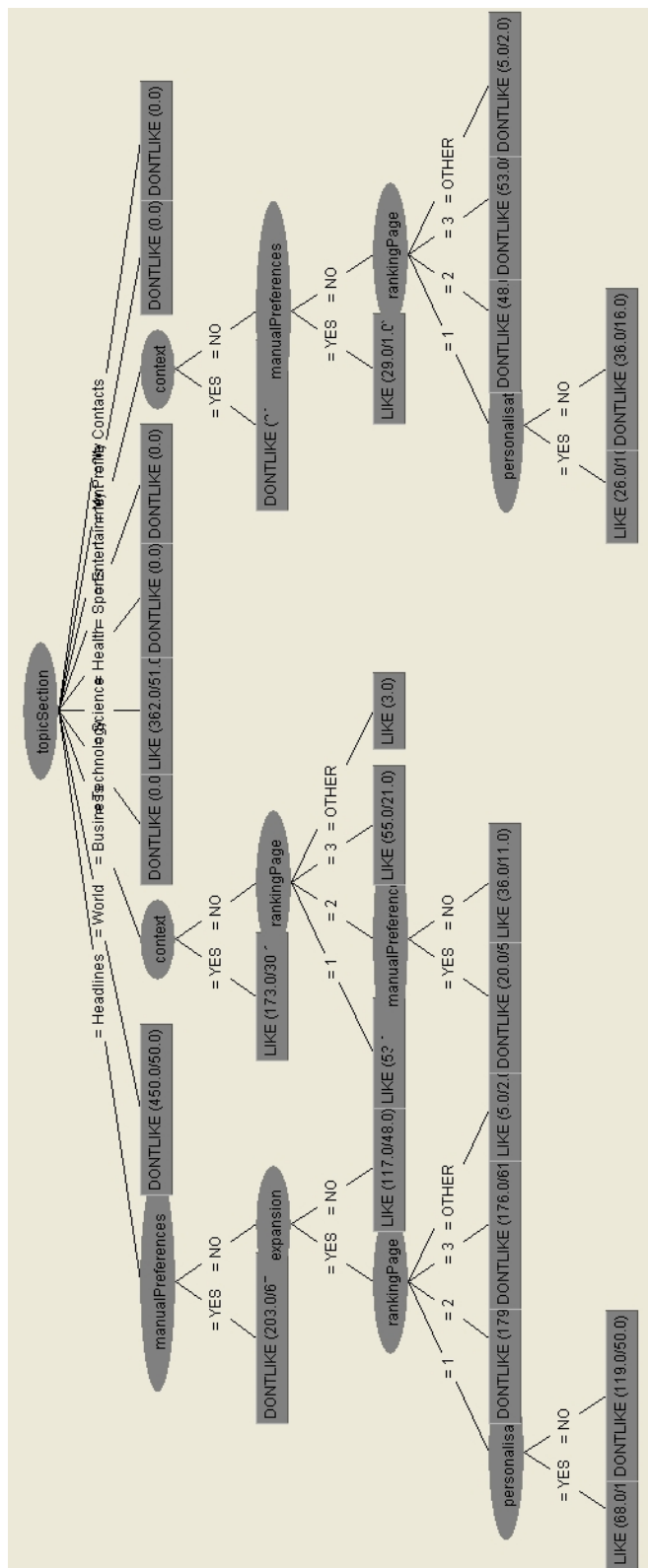


Figura 4: Árbol de decisión J48 para la evaluación general



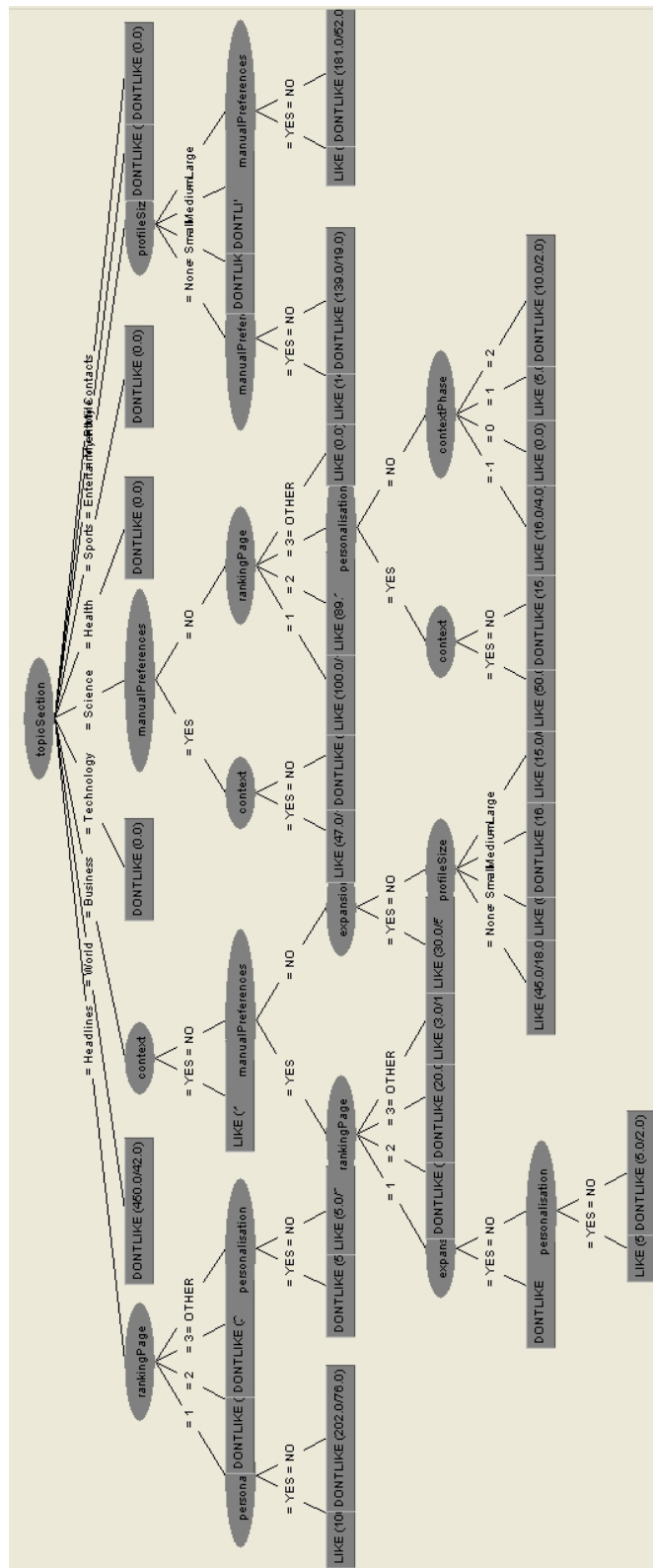


Figura 5: Árbol de decisión J48 para la evaluación orientada a perfil

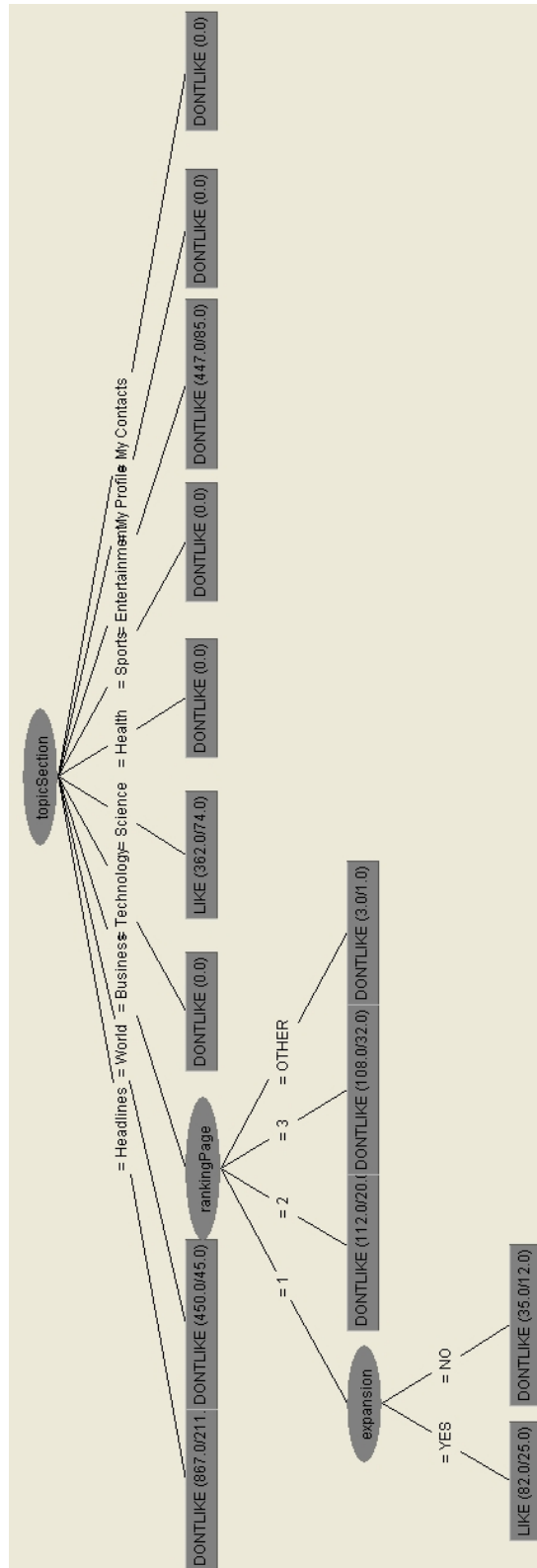


Figura 6: Árbol de decisión J48 para la evaluación general orientada a objetivo

definidas manualmente. En la sección de Ciencia, si se tienen definidas preferencias de manera manual, hay que tener activado el contexto también para encontrar noticias relevantes; en otro caso, las noticias que aparecen en las dos primeras páginas son en su mayoría relevantes con respecto al perfil, mientras que las de la tercera página depende de la activación del contexto y la personalización. Por último, en la sección de Entretenimiento la mayoría de noticias no son relevantes, excepto aquellas recuperadas con un perfil grande generado de manera manual.

- En la evaluación orientada a objetivo, hay varias secciones que, según el algoritmo de clasificación, no dependen de ningún atributo más: Portada, Internacional, Ciencia y Entretenimiento. Esto indica un claro desbalanceo a la hora de diseñar la evaluación, ya que en estas secciones, los usuarios al realizar las tareas encontraban noticias relevantes con mucha dificultad o con extremada facilidad (sección Ciencia). En la sección que queda (Economía) el conjunto de noticias que se encontraron relevantes estaban en la primera página de resultados mientras se usaba expansión.

Para poder extraer conocimiento más general, se pasa a eliminar el atributo de la sección, y se obtienen reglas a partir de árboles de decisión parciales (reglas PART) construidos usando J48. Las reglas nos permiten ver el conocimiento extraído por el algoritmo de una manera más compacta y comprensible que en el árbol. Las reglas más importantes obtenidas de esta manera se pueden ver en las tablas 2, 3 y 4.

En las reglas se puede ver el antecedente (condiciones de entrada) y el consecuente (una de las dos clases objetivo), junto con el número de ejemplos que describe correcta e incorrectamente dicha regla.

Algunos aspectos destacables que se pueden inferir de las reglas presentadas son:

- Las noticias que aparecen en la última página de resultados (tercera) generalmente no son relevantes, no obstante, esto depende de otros atributos, los cuales son distintos según el tipo de evaluación: en la evaluación general se produce en dos situaciones, la primera en la que el contexto es pequeño y no se personaliza, y la otra cuando no se realiza contexto ni personalización y las preferencias no son manuales; en la evaluación orientada a perfil, en cambio, sólo se necesita que no se active el contexto, mientras que en la evaluación orientada a objetivo se precisa que se encuentre en la última fase del contexto. Esto es importante ya que, por ejemplo en el último caso, se puede ver que las noticias no relevantes se encuentran en la última página al haber aplicado contexto dos veces, lo cual quiere decir que las noticias relevantes han aparecido con anterioridad, mostrando un comportamiento deseable.
- En la evaluación orientada a perfil se puede ver claramente cómo no aplicar personalización produce que las noticias no sean relevantes, mientras que en la evaluación orientada a objetivo esto se matiza con el hecho de que no haya contexto (fase de contexto es  $-1$ ) o que sea pequeño. En la evaluación general no se observa ninguna conclusión clara acerca de activar personalización o no.
- En la evaluación orientada a objetivo la última regla es muy significativa, ya que indica que activar contexto y utilizarlo produce resultados inmediatos (ya que en esta regla se puede ver que si se ha abierto una noticia, entonces hay un gran porcentaje de noticias relevantes).
- Aunque no existe un conocimiento claro ni determinado sobre si aplicar expansión es positivo o no (el número de ejemplos que aplica a las reglas que mencionan la expansión es muy pequeño) del caso general se puede inferir que ayuda a encontrar noticias relevantes (ya que la regla indica que se aplica a noticias que no se encontraban en la primera página).

Por último, si aplicamos una selección de atributos a cada uno de los conjuntos estudiados (sin el atributo sección), se obtiene la tabla 5. También podemos obtener los atributos ordenados por algún método de evaluación de atributos, por ejemplo, en la tabla 6 los ordenamos según la ganancia de información con respecto a la clase.

rankingPage = 3 AND  
contextSize = Small AND  
profileSize = None: DONTLIKE (212.0/71.0)

rankingPage = 1 AND  
personalisation = YES AND  
manualPreferences = NO: LIKE (304.0/79.0)

contextSize = None AND  
expansion = NO AND  
manualPreferences = YES: DONTLIKE (122.0/31.0)

contextSize = None AND  
manualPreferences = NO AND  
profileSize = Large: DONTLIKE (165.0/61.0)

personalisation = YES AND  
manualPreferences = YES: LIKE (147.0/32.0)

contextSize = Small AND  
personalisation = NO: DONTLIKE (466.0/199.0)

contextSize = None AND  
manualPreferences = NO AND  
personalisation = NO AND  
rankingPage = 3: DONTLIKE (130.0/61.0)

rankingPage = 2 AND  
personalisation = NO AND  
expansion = YES: LIKE (65.0/31.0)

rankingPage = 3 AND  
expansion = YES: LIKE (15.0/7.0)

#### Cuadro 2: Selección de las reglas extraídas de la evaluación general

rankingPage = 3 AND  
contextSize = None: DONTLIKE (320.0/84.0)

personalisation = NO: DONTLIKE (1117.0/353.0)

contextSize = None: DONTLIKE (193.0/60.0)

manualPreferences = NO: DONTLIKE (172.0/72.0)

rankingPage = 1 AND  
contextPhase = 2: LIKE (150.0/46.0)

#### Cuadro 3: Selección de las reglas extraídas de la evaluación orientada a perfil

rankingPage = 3 AND  
contextPhase = 2: DONTLIKE (326.0/56.0)

personalisation = NO AND  
rankingPage = 2: DONTLIKE (412.0/96.0)

personalisation = NO AND  
contextPhase = -1: DONTLIKE (281.0/83.0)

personalisation = NO AND  
contextSize = Small: DONTLIKE (255.0/86.0)

contextPhase = 2 AND  
manualPreferences = NO AND  
rankingPage = 1 AND  
profileSize = Large: LIKE (101.0/39.0)

contextPhase = 1: LIKE (78.0/32.0)

Cuadro 4: Selección de las reglas extraídas de la evaluación orientada a objetivo

Tipo evaluación	Atributos
General	Tamaño contexto, tamaño perfil, página de aparición
Perfil	Tamaño perfil, página de aparición
Objetivo	Expansión, tamaño perfil, página de aparición

Cuadro 5: Atributos seleccionados como más importantes para cada tipo de evaluación

Ranking	General	Perfil	Objetivo
1	Página de aparición	Página de aparición	Página de aparición
2	Tamaño perfil	Tamaño perfil	Tamaño perfil
3	Tamaño de contexto	Personalización	Personalización
4	Personalización	Preferencias manuales	Expansión
5	Preferencias manuales	Tamaño del contexto	Tamaño del contexto
6	Fase del contexto	Expansión	Fase del contexto
7	Contexto	Fase del contexto	Preferencias manuales
8	Expansión	Contexto	Contexto

Cuadro 6: Atributos ordenados por ganancia de información para cada tipo de evaluación

Si ahora se vuelve a aplicar el método de obtención de reglas PART, se encuentran como reglas positivas (la noticia es relevante: clase LIKE) las siguientes:

- Para la evaluación general:
  1. La noticia aparece en la primera página, el tamaño del contexto es pequeño y el del perfil es medio
  2. La noticia aparece en la segunda página, el tamaño del contexto es pequeño y el del perfil es grande
- Para la evaluación orientada a perfil:
  1. La noticia aparece en la primera página y el tamaño del perfil es grande
- Para la evaluación orientada a objetivo:
  1. La noticia aparece en la primera página, el tamaño del perfil es grande y hay expansión

El resto de reglas devueltas por el clasificador clasifican las noticias como no relevantes.

## 6. Conclusión

En este trabajo se ha comprobado que las técnicas de aprendizaje automático son útiles en la evaluación de entornos adaptativos. En el caso concreto de los sistemas de recomendación, donde los resultados son subjetivos por definición, estas técnicas han permitido extraer patrones comunes en determinadas situaciones, aislando el proceso de evaluación de cada usuario concreto y pudiéndose comprobar las hipótesis de partida, a saber:

- Un sistema de recomendación es más útil al usuario si utiliza personalización y contexto
- La expansión semántica favorece los resultados positivos en la recomendación

También se han podido distinguir situaciones en función del criterio de evaluación, por ejemplo, si se quiere maximizar la recomendación que más tenga en cuenta el perfil del usuario, los atributos más importantes son el tamaño del perfil y la posición del ítem, mientras que si se quiere una recomendación que tenga en cuenta el objetivo del usuario en un determinado momento, la expansión juega un papel muy importante.

Por último, hay que notar que este trabajo ha sido una primera toma de contacto con el área del aprendizaje automático para evaluar un sistema adaptativo, en el futuro se espera aplicar otras técnicas a nuevos conjuntos de datos, provenientes de evaluaciones más exhaustivas que la que utilizó este trabajo y con muchos más ejemplos a estudiar. Otro aspecto que se ha dejado para el futuro ha sido la opción de aprender qué pesos de personalización y contexto serían los más adecuados para determinadas situaciones, de manera que esta parte del sistema se pudiera adaptar de manera automática a los intereses del usuario o en función a determinados parámetros.

## 7. Agradecimientos

Quiero agradecer a Iván Cantador las ideas aportadas a este trabajo y el hacer posible que la evaluación llevada sobre el sistema News@Hand estuviera dirigida hacia la obtención de información aprovechable por los algoritmos de aprendizaje automático.

## Referencias

- [1] Jae-Wook Ahn, Peter Brusilovsky, Jonathan Grady, Daqing He, and Sue Y. Syn. Open user profiles for adaptive news systems: help or harm? In *WWW '07: Proceedings of the 16th international conference on World Wide Web*, pages 11–20, New York, NY, USA, 2007. ACM Press.
- [2] Iván Cantador, Alejandro Bellogín, and Pablo Castells. A multilayer ontology-based hybrid recommendation model. *AI Communications, special issue on Recommender Systems*.
- [3] Iván Cantador, Alejandro Bellogín, and Pablo Castells. News@hand: A semantic web approach to recommending news. In *Proceedings of the 5th International Conference on Adaptive Hypermedia (AH 2008)*. Hannover, Germany. Lecture Notes in Computer Science. Ed. Springer-Verlag, July 2008.
- [4] Iván Cantador, Miriam Fernández, and Pablo Castells. Improving ontology recommendation and reuse in webcore by collaborative assessments. In *Proceedings of the 1st International Workshop on Social and Collaborative Construction of Structured Knowledge (CKC 2007), at the 16th International World Wide Web Conference (WWW 2007)*. Banff, Canada, *CEUR Workshop Proceedings, vol. 273, ISSN 1613-0073*, May 2007.
- [5] Iván Cantador, Martin Szomszor, Harith Alani, Miriam Fernández, and Pablo Castells. Enriching ontological user profiles with tagging history for multi-domain recommendations. In *Proceedings of the 1st International Workshop on Collective Semantics: Collective Intelligence and the Semantic Web (CISWeb 2008), at the 5th European Semantic Web Conference (ESWC 2008)*. Tenerife, Spain, June 2008.
- [6] Richard O. Duda, Peter E. Hart, and David G. Stork. *Pattern Classification (2nd Edition)*. Wiley-Interscience, November 2000.
- [7] Phivos Mylonas, David Vallet, Pablo Castells, Miriam Fernandez, and Yannis Avrithis. Personalized information retrieval based on context and ontological knowledge. *Knowledge Engineering Review*, 23(1), 2008.
- [8] C. Vialardi, J. Bravo, and A. Ortigosa. Empowering aeh authors using data mining techniques. In *Fifth International Workshop on Authoring of Adaptive and Adaptable Hypermedia (A3H) held at the 11th International Conference on User Modeling (UM2007)*, June 2007.
- [9] Ian H. Witten and Eibe Frank. *Data Mining: Practical Machine Learning Tools and Techniques*. Morgan Kaufmann Series in Data Management Systems. Morgan Kaufmann, second edition, June 2005.