

Análisis de Características del Perfil de Usuario para un Sistema de Recomendación de Objetos de Aprendizaje

Paula Andrea Rodríguez^a, Néstor Darío Duque^b, Demetrio Arturo Ovalle^a, Diego Peluffo^c, Oscar Mauricio Salazar^a

^aUniversidad Nacional de Colombia – Sede Medellín, ^bUniversidad Nacional de Colombia – Sede Manizales, ^cUniversidad Cooperativa de Colombia - Pasto

Abstract. Los Objetos de Aprendizaje (OA), como unidades de contenido educativo y como recursos reutilizables permiten disponer de la información para ser utilizada en ambientes virtuales de aprendizaje y para acceder a ellos se debe contar con mecanismos de recuperación y selección permitiendo el apoyo al proceso de enseñanza-aprendizaje de un estudiante en particular. Los Sistemas de Recomendación (SR) son herramientas que apoyan al usuario a obtener información de acuerdo a sus necesidades y preferencias. Para el uso de SR en ambientes educativos, se debe contar con la caracterización de los estudiantes a través de perfiles de usuario con el fin de entregar OA adaptados. Pero no todas las características son relevantes en el proceso de recomendación, por ello en este trabajo se hace un primer acercamiento al análisis de las variables asociadas al estudiante con el fin de seleccionar aquellas que más puedan aportar al proceso de personalización.

Keywords: Reducción de dimensionalidad, características perfil usuario, objetos de aprendizaje, sistemas de recomendación.

INTRODUCCIÓN

En muchas ocasiones a los estudiantes les resulta difícil especificar completamente los términos de búsqueda para realizar una consulta y acceder acertadamente a un Objeto de Aprendizaje (OA) de su interés [1]. Si la búsqueda se realiza sólo considerando las palabras claves de la temática de interés, es muy posible que estos resultados no sean los esperados, puesto que los usuarios poseen distintas características personales y preferencias, que deberían también ser consideradas en el momento de la búsqueda [2]. Es allí donde aparecen los sistemas de recomendación, aplicaciones que pueden ofrecer sugerencias de OA a un usuario, pero para que este proceso de recomendación sea exitoso, se necesita tener información del estudiante [3]. Para representar estas características se debe definir un modelo del estudiante que incluya diversas variables en diferentes aspectos [4]. Aunque existen varios tipos de sistemas de recomendación, todos necesitan una gran cantidad de información sobre los usuarios y recursos para poder realizar recomendaciones de calidad.

En este trabajo se presenta un primer análisis estadístico de la relevancia de las características. Dicho análisis se basa en la variabilidad de los datos y está orientado a determinar la relevancia de las características en relación con los OA recomendados. Para esto se utiliza análisis de componentes principales como método de selección de características en donde se cuantifica la relevancia por cada característica. Experimentalmente se comprueba que el análisis de relevancia permite determinar efectivamente las características que conforman grupos compactos de estudiantes similares. Con esto, la recomendación puede realizarse de forma más acertada, demostrando una vez más que el estilo de aprendizaje es una variable importante al momento de hacer recomendaciones de OA.

MARCO TEÓRICO

A continuación se definen los principales conceptos en los que se enmarca este trabajo:



Perfil de Usuario

Según D'Agostino y otros un “perfil es el modelado de un objeto en forma compacta mediante sus características primordiales. En el caso de un perfil de usuario de un sistema de software, éste puede comprender tanto datos personales y características del sistema computacional, como también patrones de comportamiento, intereses personales y preferencias. Este modelo de usuario está representado por una estructura de datos adecuada para su análisis, recuperación y utilización” [5]. El uso de perfiles pretende tener información de los estudiantes con el fin de conocer sus características, necesidades y preferencias, para poder dar un resultado de una búsqueda acorde a los diferentes usuarios [2]. Entre las características que se incluye, están los datos personales, tales como, su identificación, nombre, sexo, lugar de origen, nivel de estudios, carrera, institución donde estudia; características de contexto que para este artículo se refieren a aquellas relacionadas con el tipo de dispositivo desde el cual se va a realizar la consulta; otras referentes a su historial académico donde se almacena los logros o competencias obtenidas y las actividades educativas desarrolladas con los recursos utilizados.

Aunque existen muchas variables para la caracterización de un usuario, según varios estudios, como el presentado por Alonso, Gallego y Honey en 1999, donde analizan las distintas investigaciones, concluyen que: “es suficientemente probado que los estudiantes aprenden con más efectividad cuando se les enseña en su propio Estilo de Aprendizaje”, a la vez presentan un aparte específico de los diferentes modelos para la selección de un estilo de aprendizaje. El término estilos de aprendizaje (Learning Styles, en inglés) se refiere a las estrategias preferidas de recopilar, interpretar, organizar y pensar sobre la nueva información [6]. Existen varios modelos para representar los estilos de aprendizaje como [7]: Modelo de Felder y Silverman: está diseñado a partir de cuatro escalas bipolares que son Activo-Reflexivo, Sensorial-Intuitivo, Visual-Verbal y Secuencial-Global; Modelo de Kolb: supone que para aprender algo se debe trabajar o procesar la información que se recibe, definiendo entonces alumno activo, teórico, reflexivo o pragmático; Modelo VARK: definen las preferencias al procesar la información desde el punto de vista sensorial (Visual, Auditivo, Lecto-escritor y Kinestésico).

En [7] se presenta la combinación de los modelos VARK y FSLSM obteniendo buenos resultados para caracterizar el perfil del estudiante y de esta manera entregar al estudiante los materiales educativos adaptados.

Sistemas de Recomendación

Los Sistemas de Recomendación surgieron a mediados de la década de los 90 [8] con el fin de brindarle a los usuarios aquellos resultados de búsquedas de información cercanos a sus necesidades [2] [9]. Chesani (2007) define un sistema de recomendación como aquel capaz de realizar predicciones a partir del hecho que a un usuario le guste o no cierto ítem, objeto, medio o información al que podría acceder [10],[11]. Los sistemas de recomendación son los aliados de la personalización de sistemas computacionales, principalmente en la web, por su capacidad de identificar preferencias y sugerir ítems relevantes para cada usuario [4]. Existen muchos tipos de sistemas de recomendación, cada uno puede ser usado para diferentes enfoques y en diferentes contextos, de acuerdo a las necesidades [12].

Análisis Estadístico de Características

En este trabajo se hace un análisis de relevancia de las características para determinar cuáles de ellas son las que mejor representan los grupos de estudiantes con perfiles similares. En particular, en este estudio se realiza dicho análisis a partir de la variabilidad de las características usando un método de estadística convencional denominado Análisis de Componentes Principales (PCA –de sus siglas en inglés *Principal Component Analysis*) [13]. El método clásico de PCA se aplica sobre datos cuantitativos, por tanto las características consideradas deben ser estrictamente numéricas y legítimamente cuantitativas.

Suponga que se tienen n estudiantes, cada uno representado con d características. Estos datos podrían organizarse en una matriz X , de forma que sea de dimensiones $n \times d$. Las columnas de dicha matriz representan los

ejes coordenados del espacio original de representación. El objetivo de PCA es encontrar un nuevo espacio de representación en donde la mayor varianza es acumulada en los primeros ejes o componentes principales [13]. Los componentes principales son ejes de representación para el nuevo espacio y son obtenidos de un análisis de la matriz de covarianza S que en este caso es de dimensiones $d \times d$. La matriz S se puede calcular generalizando el concepto de varianza a casos multivariados. Una vez se centra la matriz X asegurando que cada columna tenga media igual a cero, la matriz de covarianza se puede calcular como:

$$S = \frac{1}{n} X^T X, \quad (1)$$

donde X^T denota la matriz transpuesta de X . En el trabajo realizado en [14], se demuestra que la relevancia de las características puede calcularse con una suma ponderada de los vectores propios de matriz de covarianza. También se demuestra que los factores de ponderación de dicha suma son los respectivos valores propios. Dado que S es de dimensión d , se tienen d valores propios que se denotan como $\lambda_1, \dots, \lambda_d$ ordenados de forma que $\lambda_1 > \lambda_2 > \dots > \lambda_d$. Los correspondientes vectores propios se denotan con v_1, \dots, v_d . La relevancia de una característica en particular, denotada con c , tal que $c=1, \dots, d$, está dada por [13]:

$$rel(c) = \sum_{i=1}^p \lambda_i v_i(c)^2, \quad (2)$$

donde $v_i(c)$ es la posición c es del vector i -ésimo. La variable p se calcula con criterio de varianza acumulada menor del 100%. Esta variable es tal que $p < d$ y significa que se usan los componentes principales que más aportan a la varianza tal como se explica en [14].

DESCRIPCIÓN DE EXPERIMENTOS

Se seleccionaron 34 estudiantes de Administración de Sistemas Informáticos de la Universidad Nacional de Colombia Sede Manizales, de semestres entre quinto y decimo, a los cuales se les pidió que llenaran un formulario de registro, donde se pedía información personal (nombres, fecha de nacimiento, sexo), preferencias respecto a un tema y finalmente se pidió que realizaran el test de estilo de aprendizaje para determinar como el estudiante procesa la información. Para el test, se tomaron las 13 preguntas del test de VARK [15] y las preguntas relacionadas con la dicotomía secuencial/global del test de Felder Silverman [16]. Se seleccionaron estos test, debido a trabajos previos del grupo de investigación, donde se encontraron buenos resultados [7].

Para hacer las validaciones a cada uno de los estudiantes se le recomendaron en un conjunto de 10 objetos de aprendizaje cuyos ID son los números entre 100 y 109 (ver Tabla 1), relacionados con la temática de programación, según la técnica de recomendación propuesta en [17] donde se toma el estilo de aprendizaje como la característica más importante. Este proceso de recomendación consta de tres técnicas diferentes, la primera de ellas es la recomendación por contenido donde se seleccionan los OA a recomendar según el estilo de aprendizaje que tiene el estudiante al que se le entregarán las recomendaciones, el segundo es basado en filtrado colaborativo donde el resultado de la recomendación son los OA que evaluaron positivamente, en una escala de 1 a 5, donde 5 es muy relevante y 1 no es nada relevante, estudiantes con perfiles similares al estudiante al que se le está recomendando, y finalmente la tercera técnica es según conocimiento, donde los resultados de la recomendación son los OA similares a los que, el estudiante objetivo de la recomendación, ya evaluó positivamente en el pasado. Finalmente, los resultados de las tres técnicas de recomendación se integran en lo que se llama un SR híbrido, seleccionando las intersecciones de los OA resultados para entregar los OA que fueron recomendados por las tres técnicas [17].

Para aplicar análisis de relevancia de características a través de PCA, se requiere estrictamente de datos cuantitativos, por tanto las características consideradas, con la información disponible, fueron edad, sexo, resultado test VARK y resultado test de Felder Silverman. La edad se calculó con la fecha de nacimiento del estudiante, el sexo se definió de forma que 0 es mujer y 1 es hombre; para el test de VARK: 1 visual, 2 auditivo, 3 lector y 4 kinestésico y finalmente se hizo una simplificación para el test de Felder Silverman definiendo 0 para

tendencia global y 1 para secuencial. En la Tabla 1 se aprecia un fragmento de la matriz de datos considerados que presentan los 34 estudiantes con los datos que definen su perfil y los ID de los 3 conjuntos de OA recomendados previamente.

TABLA (1). Matriz de datos y ID de los conjuntos de objetos recomendados

| # | Edad | Sexo | VARK | Felder | OA1 | OA2 | OA3 |
|-----|------|------|------|--------|-----|-----|-----|
| 1 | 21 | 1 | 1 | 1 | 102 | 103 | 106 |
| 2 | 23 | 0 | 2 | 0 | 104 | 105 | 105 |
| 3 | 30 | 1 | 3 | 0 | 107 | 107 | 107 |
| 4 | 21 | 1 | 1 | 1 | 102 | 103 | 106 |
| ... | ... | ... | ... | ... | ... | ... | ... |
| 31 | 69 | 1 | 1 | 1 | 102 | 103 | 106 |
| 32 | 44 | 1 | 1 | 1 | 102 | 103 | 106 |
| 33 | 28 | 0 | 1 | 1 | 102 | 103 | 106 |
| 34 | 27 | 0 | 3 | 0 | 107 | 108 | 108 |

Se tiene que la matriz tiene $n=34$ filas y $d=4$ columnas. Los experimentos consisten en aplicar PCA a estas características y se calculan los valores de relevancia a partir de la ecuación (2) con los datos y vectores propios de la covarianza de la ecuación (1). Para observar la relación entre estudiantes con perfiles similares y objetos recomendados, se realizan diagramas de dispersión o *scatter plots* de la matriz de datos con respecto a los OA recomendados, es decir, imprimiendo de diferente color los puntos en el espacio 2D o 3D de acuerdo con el OA asignado como se aprecia en la Figura 2. Posteriormente, se realizan observaciones sobre los gráficos obtenidos para comprobar la pertinencia del análisis de relevancia.

RESULTADOS Y DISCUSIÓN

En la Figura 1a) se presenta la matriz de los datos originales, cada color representa un valor dentro del rango de su columna, en las columnas esta cada atributo (características) y en las filas esta cada una de las instancias de un estudiante. En la Figura 1b) se puede observar el resultado de PCA, se grafica $rel(c)$ vs c , para observar los valores de relevancia de cada característica. Se aprecia que las características más relevantes son las 3 y 4, concluyendo que los componentes principales en este estudio son el estilo de aprendizaje, representado en dos test, característica 3 test de VARK y característica 4 test de Felder Silverman, seguido de la edad y por último el sexo como el componente menos relevante. Con lo que se demuestra que el estilo de aprendizaje es una variable importante al momento de hacer recomendaciones de OA.

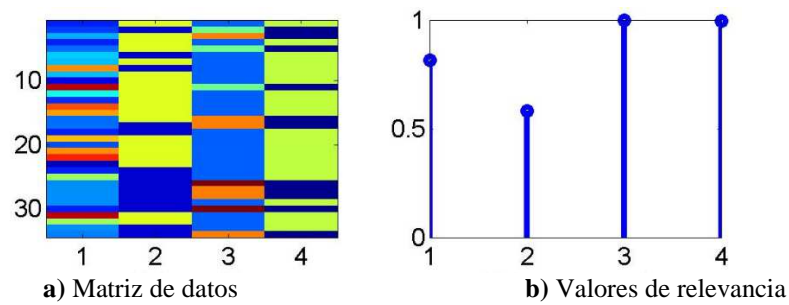
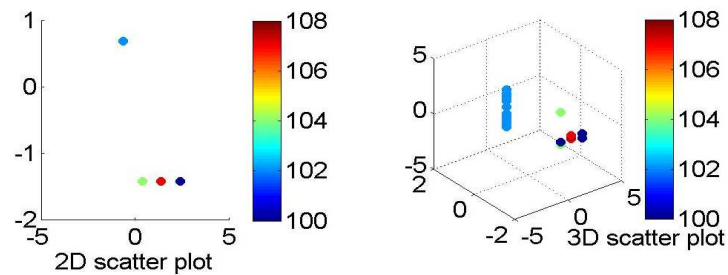
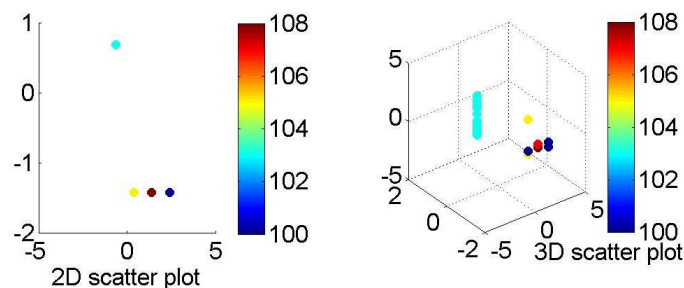


Figura 1. a) Representación gráfica de la matriz de datos: cada número es representado con un color diferente. Esto permite notar visualmente que variables definen grupos similares. **b)** Valores de relevancia de las 4 características consideradas, el eje vertical es el valor de relevancia y el eje horizontal es la correspondencia con las características.

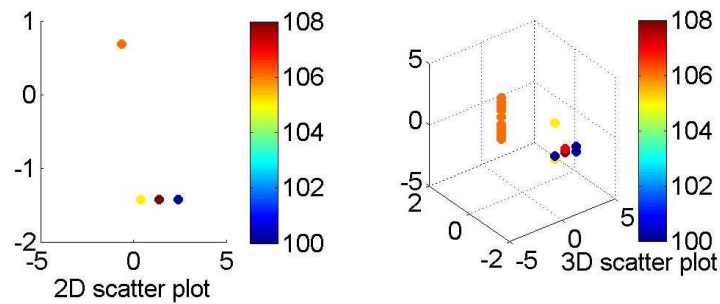
Se hicieron análisis con los OA recomendados, a cada estudiante se le recomendó un máximo de 3 OA, por lo tanto en la Figura 2 se presentan cómo se agrupó cada estudiante y con los colores se representa los OA recomendados, en otras palabras cada punto es un estudiante, el color de ese punto representa el OA que se le recomendó a ese estudiante en el conjunto de datos 1, Figura 2a), conjunto 2 en la Figura 2b) y el conjunto 3 en la Figura 2c).



a) Primer conjunto de objetos recomendados



b) Segundo conjunto de objetos recomendados



c) Tercer conjunto de objetos recomendados

FIGURA 2. Diagramas de dispersión o *scatter plots* de los características relevantes en 2D y 3D para los tres conjuntos de objetos recomendados.

Se aprecia que al usar las variables relevantes los grupos formados de estudiante son consistentes (es decir que los estudiantes a quienes se les recomendó el mismo objeto se ven cercanos en el espacio) y relativamente compactos. En otras palabras, los OA son representados a través de los *scatter plots* de los estudiantes. Por tanto, si se ven puntos del espacio (estudiantes) del mismo color cercanos es decir, formando grupos compactos y sin traslapes (puntos de otro color), la recomendación es consistente con las características. Este comportamiento se observó con los tres conjuntos de OA recomendados se ve un comportamiento similar. El scatter 2D representa las dos características más relevantes y se nota que los grupos son muy compactos, aspecto que también se puede apreciar en la representación gráfica de la matriz en la Figura 1 o bien en la matriz de la Tabla I. Por otra parte, cuando se utiliza la variable menos relevante, el sexo en este caso, con la segunda menos relevante, se aprecia que no se forman grupos claros ni compactos, como puede verse en la Figura 3. Dado que estudiantes a quienes se les recomendó el mismo objeto u similar deben verse cercanos en el *scatter* de las características, se puede concluir que un análisis de relevancia permite determinar las características más adecuadas para hacer recomendaciones acertadas y consistentes.

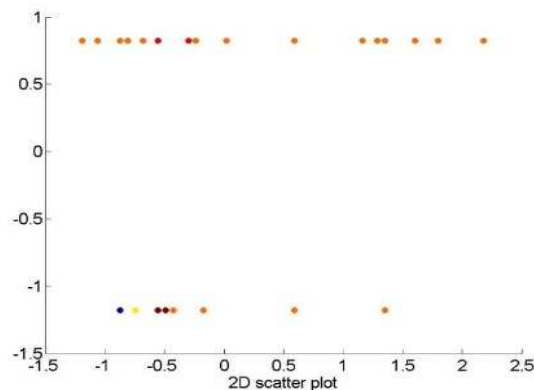


FIGURA 3. Diagramas de dispersión o *scatter plots* de los características menos relevantes (o no relevantes) en 2D. Este diagrama es un ejemplo seleccionando uno de los conjuntos de objetos recomendados. Se aprecia que no se forman grupos homogéneos con estas características y por tanto no se puede hacer una buena recomendación.

CONCLUSIONES Y TRABAJO FUTURO

En este artículo se realiza un primer acercamiento al análisis de relevancia de las características disponibles en un perfil de usuario – estudiante, que sirvan como insumo en un sistema de recomendación de objetos de aprendizaje para entregar materiales adaptados a su especificidad. Los OA recomendados pueden ser usados como recursos de formación a los estudiantes apoyando su proceso de enseñanza – aprendizaje.

Como trabajo futuro se plantea realizar de nuevo el experimento teniendo en cuenta más características del perfil del estudiante, contar con más cantidad de perfiles y tomar un conjunto de registros más amplio.

ACKNOWLEDGMENTS

El trabajo de investigación presentado en este artículo fue financiado parcialmente por el proyecto de COLCIENCIAS titulado: "RAIM: Implementación de un framework apoyado en tecnologías móviles y de realidad aumentada para entornos educativos ubicuos, adaptativos, accesibles e interactivos para todos" de la Universidad Nacional de Colombia, con código 1119-569-34172.

Este trabajo fue desarrollado con el apoyo de la beca otorgada por "Programa Nacional de Formación de Investigadores – COLCIENCIAS".

REFERENCIAS

1. T. I. Wang, K. H. Tsai, M. C. Lee, and T. K. Chiu, "Personalized Learning Objects Recommendation based on the Semantic Aware Discovery and the Learner Preference Pattern," *Educ. Technol. Soc.*, vol. 10, no. 3, pp. 84–105, 2007.
2. A. Casali, V. Gerling, C. Deco, and C. Bender, "Sistema inteligente para la recomendación de objetos de aprendizaje," *Rev. Generación Digit.*, vol. 9, no. 1, pp. 88–95, 2011.
3. O. Sanjuán, E. Torres, H. Castán, R. Gonzalez, C. Pelayo, and L. Rodriguez, "Viabilidad de la aplicación de Sistemas de Recomendación a entornos de e-learning," *Univ. Oviedo, España*, 2009.
4. S. C. Cazella, E. B. Reategui, and M. A. Nunes, "A Ciência da Opinião: Estado da arte em Sistemas de Recomendação," *JAI Jorn. Atualização em Informática da SBC. Rio Janeiro, RJ PUC Rio*, pp. 161–216, 2010.
5. E. DAgostino, A. Casali, R. Corti, and A. Torres, "Sistema de Apoyo al Aprendizaje Diagnóstico Utilizando Perfiles de Usuario : EndoDiag II," *eia.udg.es*, pp. 1–14, 2005.
6. C. Alonso, D. Gallego, and P. Honey, *Los Estilos de Aprendizaje. Procedimientos de diagnostico y mejora*. Bilbao, 1997.
7. H. González, N. Duque Méndez, and D. Ovalle C., "Técnicas inteligentes para la actualización dinámica del perfil del usuario en un sistema de educación virtual," in *Tendencias en Ingeniería de Software e Inteligencia Artificial*, 2009.
8. J. Z. Li, "Quality, Evaluation and Recommendation for Learning Object," *Int. Conf. Educ. Inf. Technol.*, no. Iceit, pp. 533–537, 2010.
9. R. Burke, "Hybrid Recommender Systems: Survey and Experiments," *User Model. User-adapt. Interact.*, vol. 12, no. 4, pp. 331–370, 2002.
10. F. Chesani, "Recommendation Systems," *Corso di laurea Ing. Inform.*, pp. 1–32., 2002.
11. K. Mizhquero, "Análisis , Diseño e Implementación de un Sistema Adaptivo de Recomendación de Información Basado en Mashups," *Rev. Tecnológica ESPOL*, 2009.
12. F. Hdioud, B. Frikh, and B. Ouhbi, "A comparison study of some algorithms in Recommender Systems," *Inf. Sci. Technol.*, vol. Colloquium, pp. 130–135, 2012.
13. D. H. Peluffo-Ordóñez, J. L. Rodríguez-Sotelo, and G. Castellanos-Domínguez, "Estudio comparativo de métodos de selección de características de inferencia supervisada y no supervisada," *Revista Tecnológicas*, no. 23, pp. 149–166, 2009.
14. D. H. Peluffo, J. A. Lee, M. Verleysen, J. L. Rodríguez, and G. Castellanos, "Unsupervised relevance analysis for feature extraction and selection A distance-based approach for feature relevance," in *International conference on pattern recognition, applications and methods - ICPRAM 2014*, 2014.
15. N. Fleming and D. Baume, "Learning Styles Again: VARKing up the right tree!," *Educ. Dev.*, no. 7, 2006.
16. R. Felder and L. Silverman, "Learning and teaching styles in engineering education," *Eng. Educ.*, vol. 78, no. June, pp. 674–681, 1988.
17. P. Rodriguez, V. Tabares, D. Ovalle, and N. Duque, "Modelo multi-agente para recomendación híbrida de objetos de aprendizaje," *Rev. virtual Univ. Católica del Norte*, vol. 40, pp. 96–110, 2013.