

Explotación del B2B y segmentación de la clientela

Algoritmos de recomendación en R

Sesión 4. 24/10/2018

A.M.Mayoral (asun.mayoral@umh.es), J.Morales (j.morales@umh.es)

Explotación del B2B y segmentación de la clientela

Se presentarán las posibilidades de las técnicas de segmentación y cluster, así como los fundamentos de los algoritmos de recomendación. El tratamiento de la información disponible y proporcionada por los clientes permitirá a la empresa diseñar nuevas estrategias de marketing y ventas. Estas técnicas se ilustrarán sobre bases de datos reales en temáticas susceptibles de ser adaptadas en el ámbito del calzado.

Sistemas de recomendación

Un sistema de recomendación aplica algoritmos basados en técnicas estadísticas al problema de proporcionar, de un modo automatizado, recomendaciones de productos a partir de datos previos relativos a las elecciones de los usuarios e información de los productos.

- Programación
- Matemáticas
- Estadística
- Tecnología

Impacto económico relevante:
Amazon, Netflix (Netflix Prize, 2006-09), eBay, ...

Elementos de un sistema de recomendación



Medidas de valoración

Binaria



Ordinal



Numérica

visualizaciones
clicks

...

Medidas de caracterización: atributos

Binaria



Ordinal



Numérica

visualizaciones
clicks
...

Metadatos: Presencia / Ausencia



Medidas de distancia. Datos numéricos

Distancia euclídea

raíz de la suma de las distancias al cuadrado en todas las dimensiones

Distancia de Manhattan

suma de las distancias en todas las dimensiones

Correlación

(Pearson, Spearman, Kendall,...)
distancias al centro estandarizadas

Correlación Jackknife

correlación prevenida de outliers
1-CR

Distancia coseno

orientaciones similares
1-DC

estandarización de valoraciones

similitud por patrón para evitar el efecto escala

Medidas de distancia. Datos numéricos

Euclídea

$$\sqrt{\sum_i^n (x_i - y_i)^2}$$

Manhattan

$$\sum_{i=1}^n |x_i - y_i|$$

Correlación

$$\frac{\sum (x - \bar{x})(y - \bar{y})}{\sqrt{\sum (x - \bar{x})^2 \sum (y - \bar{y})^2}}$$

Jackknife

min. correlaciones sin una observación

Coseno

$$\sum_{i=1}^n \frac{x_i y_i}{\sqrt{x_i^2} \sqrt{y_i^2}}$$

estandarización
de valoraciones

similitud por patrón para
evitar el efecto escala

Medidas de distancia. Datos binarios

Simple
matching coeff

$$(b+c)/(a+b+c+d)$$

Índice Jaccard

$$(b+c)/(a+b+c)$$

...

Más en

- Cha, Choi & Tappert (2010)
- Jossin and Sulaiman (2005)

	1	0
1	a	b
0	c	d

Sistema de recomendación

Set de usuarios u_1, u_2, \dots, u_m
Set de items i_1, i_2, \dots, i_n

Objetivo: recomendar S items no valorados por el usuario u_x

R = matriz de valoración $m \times n$

- usuarios en filas
- items en columnas

No todos los usuarios tienen por qué haber hecho una valoración de todos los items.

A partir de las predicciones de los items no valorados se seleccionan los top S .

Sistemas de recomendación

Basados en
contenido
(content based filtering)

muéstrame más
cosas como las
que me han
gustado

Filtrado
Colaborativo
(collaborative filtering)

muéstrame cosas
como las que le
han gustado a
gente parecida a
mí

Mixtos
(mixed filtering)

muéstrame cosas
como las que me
han gustado a mí
o a gente parecida
a mí

Sistemas de recomendación

Basados en contenido

(ej: música)

muéstrame más cosas como las que me han gustado

- atributos en los productos
- valoraciones de cada sujeto

Filtrado Colaborativo

(ej: Amazon)

muéstrame cosas que le han gustado a gente parecida a mí

- valoraciones de sujetos+info perfil

basado en el usuario
(UserBased)

basado en el ítem
(ItemBased)

Mixtos

muéstrame cosas como las que me han gustado a mí o a gente parecida a mí

Filtrado basado en contenidos

Set de usuarios u_1, u_2, \dots, u_m
Set de items i_1, i_2, \dots, i_n

Objetivo: recomendar S items no valorados por el usuario u_x en base a atributos y la experiencia del usuario u_x

1. Cálculo de similitud (respecto atributos) de todos los items no valorados con los items valorados por el usuario u_x .
2. Selección de los N items (valorados) más afines a cada uno de los no valorados (N via validación).
3. Cálculo del índice de recomendación (predicción) con el promedio (ponderado por similitud) de las valoraciones del usuario u_x a los N items más afines.
4. Selección de los S items con mayor índice de recomendación.

Filtrado colaborativo basado en items

Set de usuarios u_1, u_2, \dots, u_m
Set de items i_1, i_2, \dots, i_n

Objetivo: recomendar S items no valorados por el usuario u_x con valoraciones de otros items y las del usuario u_x .

1. Identificación de items “relevantes”.
2. Cálculo de similitud de todos los items no valorados con los items valorados por el usuario u_x (correlaciones con valoraciones de todos los usuarios).
3. Selección de los N items valorados más afines (correlación positiva) para cada uno de los no valorados.
4. Cálculo del índice de recomendación (predicción) con el promedio (ponderado por similitud) de las valoraciones del usuario u_x a los N items más afines.
5. Selección de los S items con mayor índice de recomendación.

Filtrado colaborativo basado en usuarios

Set de usuarios u_1, u_2, \dots, u_m
Set de items i_1, i_2, \dots, i_n

Objetivo: recomendar S items no valorados por el usuario u_x con valoraciones de usuarios afines.

1. Identificación de usuarios “activos”.
2. Cálculo de similitud entre todos los usuarios, en base a sus valoraciones (correlaciones).
3. Selección de los N usuarios más afines a u_x (N via validación).
4. Cálculo del índice de recomendación (predicción) con el promedio (ponderado por similitud) de las valoraciones medias (si suficientes) de los N usuarios afines.
5. Selección de los S items con mayor índice de recomendación.

Posible eliminación de usuarios con varianza muy pequeña (no discriminan) / ponderación por la varianza.

Sistema de recomendación.

Ejemplos

(recommenderlab)

MovieLense

Valoraciones de 943
usuarios sobre 1664
películas catalogadas
respecto de 19 géneros

Objetivo:

recomendar películas al
usuario $x=329$

Sistema de recomendación.

Ejemplos

(recommenderlab)

Jester5k

Valoraciones de chistes de 5000 usuarios del sistema de recomendación Jester Online Joke. Recopilados entre abril 1999 y mayo 2003.

<http://eigentaste.berkeley.edu/index.html>

Objetivo:

recomendar chistes al
usuario $x=329$

Apache Mahout	Machine learning library, includes collaborative filtering	Java	http://mahout.apache.org
Cofi	Collaborativ filtering library	Java	http://www.nongnu.org/cofi/
Crab	Components to create recom	Python	https://github.com/muricoca/crab
easyrec	Recommender for Web pages	Java	http://easyrec.org/
LensKit	Collaborative filtering algorithms from GroupLens Research	Java	http://lenskit.grouplens.org/
MyMediaLite	Recommender system algorithms	C#/Mono	http://mloss.org/sorgware/view/282/
PREA	Personalized Recommendation Algorithms Toolkit	Java	http://prea.gatecg.edu/
SVDFeature	Toolkit for feature-based matrix facORIZATION	C++	http://mloss.org/software/view/333/
Vogoo PHP LIB	Collaborative filtering engine for personalizing web sites	PHP	http://sourceforge.net/projects/vogo o

Referencias

Choi, S.S. Cha, S.H. and Tappert,C.C. (2010). A survey of Binary Similarity and Distance measures. Systemics, Cybernetics and Informatics, Vol.8,No1., 43-48.

[.http://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.352.6123](http://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.352.6123)

Hossin, M. and Sulaiman, M.N. (2005). A review on Evaluation Metrics for Data Classification Evaluations. International Journal of Data Mining & Knowledge Management Process, Vol.5, No.2.

<https://www.ijecs.in/index.php/ijecs/article/download/1137/1035/>

Malone TW, Grant KR, Turbak FA, Brobst SA, Cohen MD (1987). Intelligent information-sharing systems. Communications of the ACM, 30(5), 390{402. <http://doi.acm.org/10.1145/22899.22903>.

Park D.H., Kim, H.K., Choi, I.L., Kim, J.K. (2012). A literature review and classification of recommender systems research. Expert Systems with Applications

Volume 39,11, pg. 10059-10072

<https://doi.org/10.1016/j.eswa.2012.02.038>

Referencias

Linden, G., Smith, B., York, J. (2003). Amazon.com recommendations. Item-to-item collaborative filtering. Recuperado en octubre 2018
<http://www.cs.umd.edu/~samir/498/Amazon-Recommendations.pdf>

Ricci F., Rokach L., Shapira B. (2015) Recommender Systems: Introduction and Challenges. In: Ricci F., Rokach L., Shapira B. (eds) Recommender Systems Handbook. Springer, Boston, MA. https://doi.org/10.1007/978-1-4899-7637-6_1

Gomez-Uribe, C.A., Hunt, N. (2015). The Netflix Recommender System: Algorithms, Business Value, and Innovation. ACM Transactions on Management Information Systems, Vol.6, No4, Article 13. Recuperado en octubre 2018
<https://dl.acm.org/citation.cfm?id=2843948>.

Ansari A, Essegaiier S, Kohli R (2000). Internet Recommendation Systems." Journal of Marketing Research, 37, 363-375.
<https://doi.org/10.1509/jmkr.37.3.363.18779>

Referencias

Hahsler, M. (2015). recommenderlab: A Framework for Developing and Testing Recommendation Algorithms. Recuperado en oct2018 <http://bit.ly/2z6VaZB>.

Amat Rodrigo, J. (2018). Sistemas de recomendación con R.
https://rpubs.com/Joaquin_AR/370301