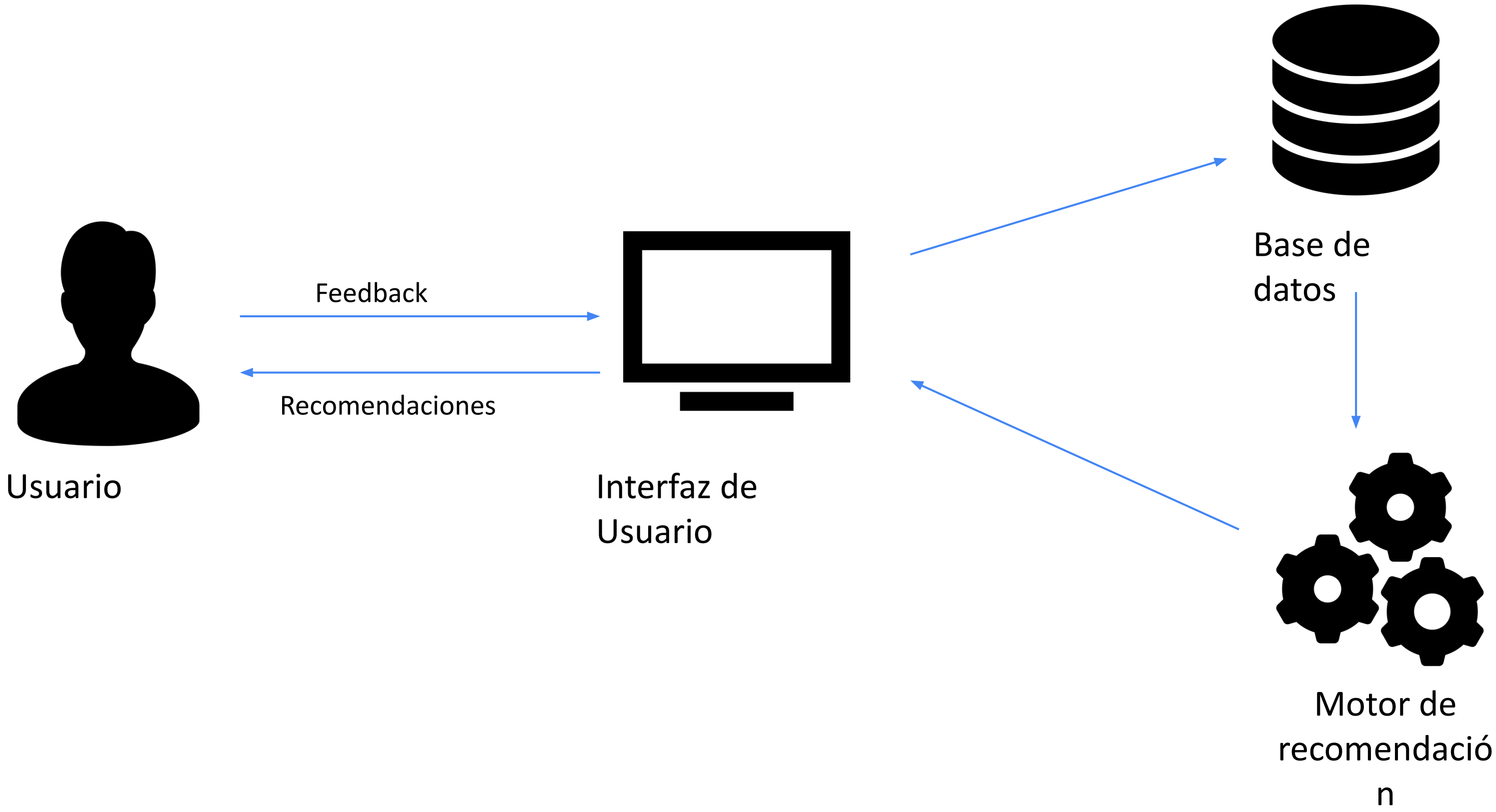
Sistemas Recomendadores IIC-3633

Repaso general y
Trabajo Futuro en Sistemas Recomendadores

Repaso general

Esquema de recomendación



Recomendación no personalizada, semi-personalizada y personalizada.

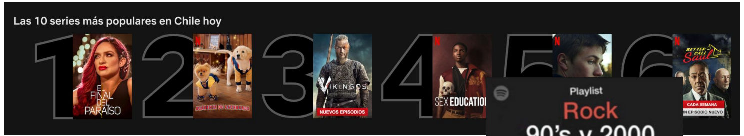
No personalizada

Semi- personalizada

Personalizada

- Más populares
- Random
- L2R

Basada en regla general (ej País, Segmentación por edad, intereses)

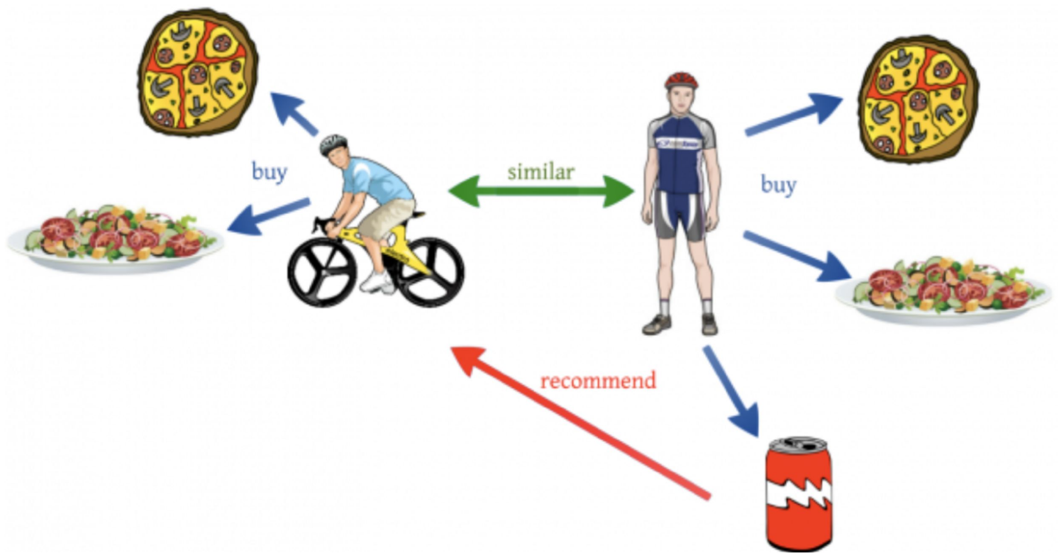


Basada en intereses personales

Filtrado colaborativo
Factores latentes
Basada en contenido
Ensambles / Híbridos

Filtrado Colaborativo basado en Usuarios

- **Objetivo:** buscar a usuarios similares y recomendar usando una suma ponderada con una métrica de similaridad



https://www.slideshare.net/tantrieuf31/introduction-to-recommendation-systems

Filtrado Colaborativo basado en Items

- En vez de calcular la similaridad entre usuarios, calcula la similaridad entre ítems para generar las recomendaciones
- Sub tareas
  - Calcular similaridad entre ítems *co-rated* (co-consumidos)
  - Calcular predicciones

USER-BASED

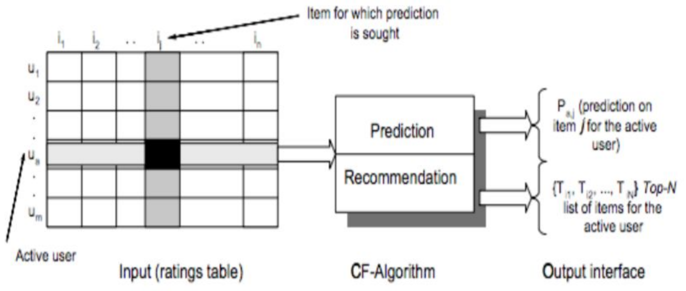


Figure 1: The Collaborative Filtering Process.

ITEM-BASED

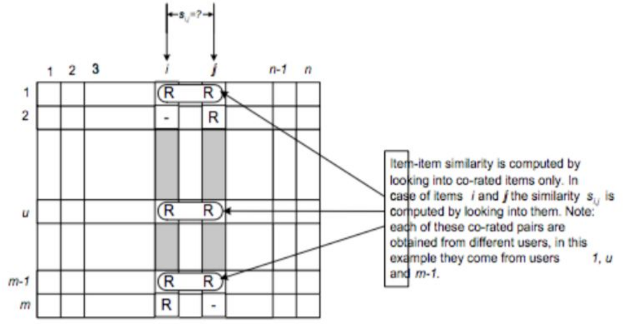
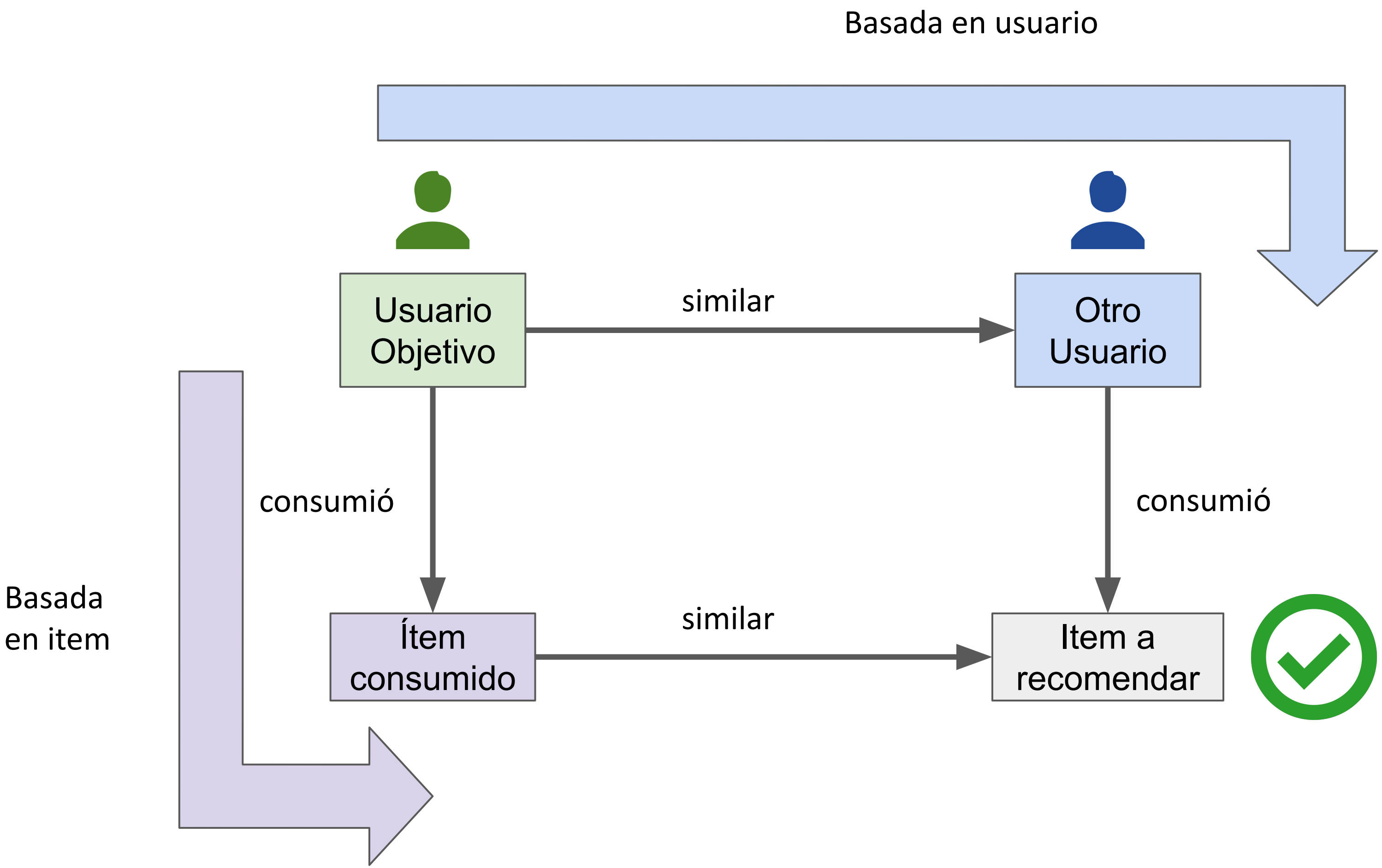
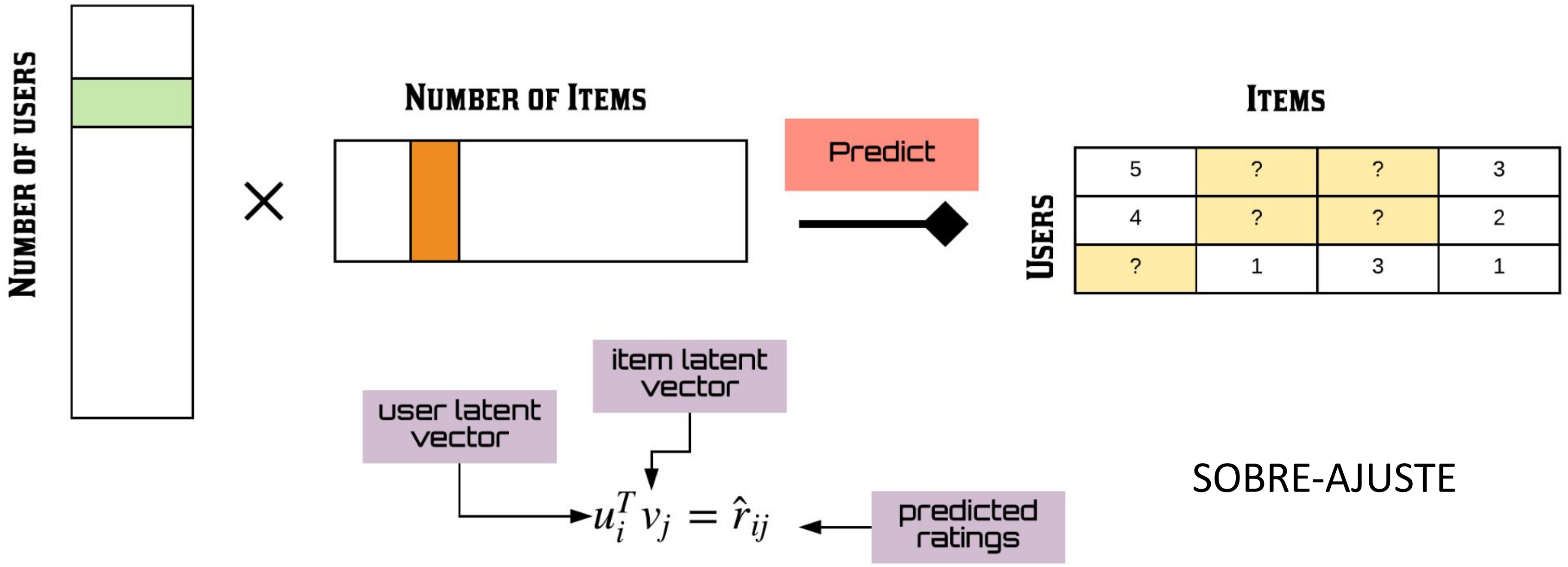


Figure 2: Isolation of the co-rated items and similarity computation

Filtrado colaborativo (resumen)



Recomendación basada en factores latentes (factorización matricial SVD).



SOBRE-AJUSTE

Feedback implícito

Retroalimentación Implícita o Implicit Feedback

- Hu, Y., Koren, Y., & Volinsky, C. (2008) modificaron el modelo de SVD para incluir *feedback* implícito

Binarización de los datos

$$p_{ui} = \begin{cases} 1 \text{ si } r_{ui} > 0 \\ 0 \text{ si } r_{ui} = 0 \end{cases}$$

Ej. lo compra o no lo compra?
interacción?

Función de confianza

$$c_{ui} = 1 + \alpha r_{ui}$$

Ej. - dwell time (más tiempo , mayor score)
- monto $, UF

Función de pérdida (Implicit Funk SVD)

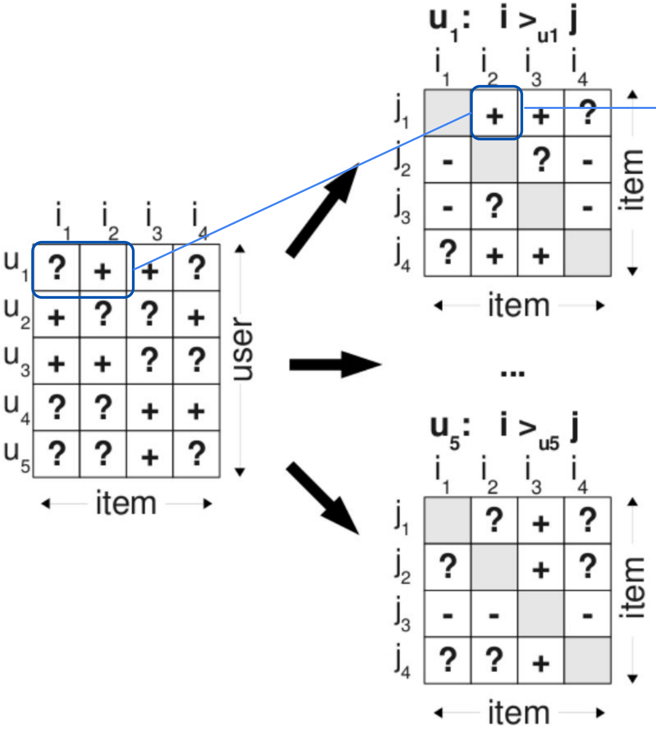
$$\min_{x^*, y^*} \sum_{u,i} c_{ui}(p_{ui} - x_u^T y_i)^2 + \lambda(||x_u||^2 + ||y_i||^2)$$

Alternativas de FM para feedback implícito

Alternating Least Squares (ALS): aprende vector latente de usuarios e ítems de manera alternada.

Bayesian Personalized Ranking (BPR): aprender a rankear ítems relevantes en las primeras posiciones comparando pares.

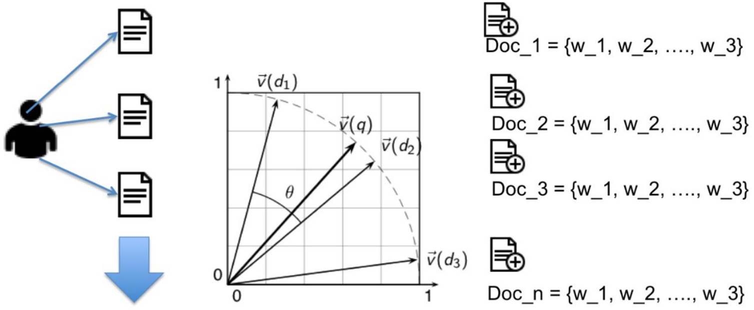
Transformando retroalimentación en positiva y negativa



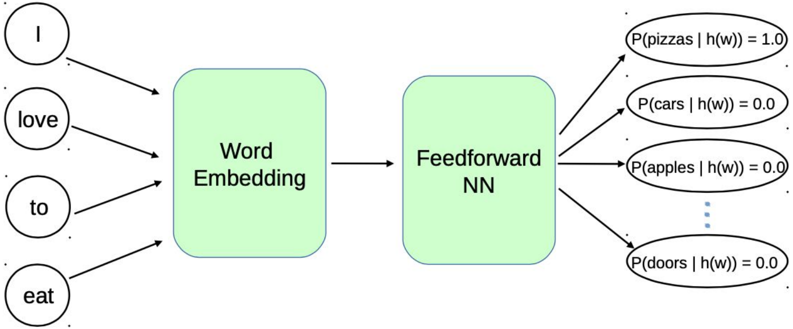
$$(u_1, i_2, j_1)$$

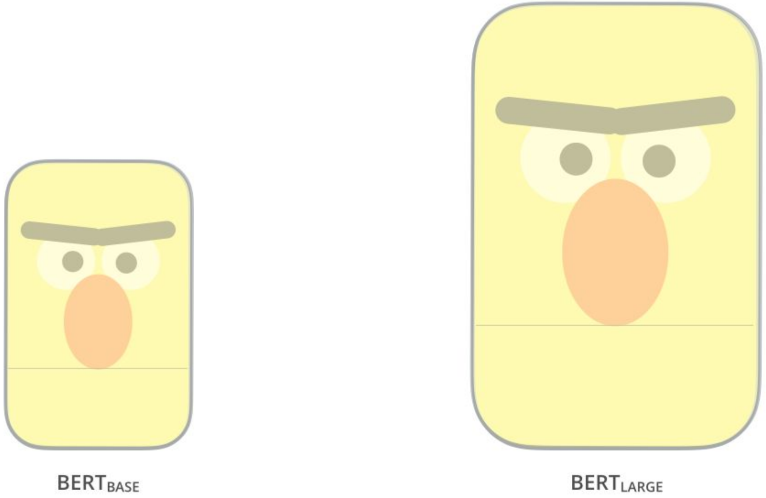
Cada usuario (ui) se representa como una matriz de preferencias.

$$\mathcal{D}_p = \{(u, i, j) \in \mathcal{D} | i \in \mathcal{I}_u \wedge j \in \mathcal{I} \setminus \mathcal{I}_u\}$$



user_profile = {w_1, w_2, …., w_3} usando
TF-IDF





Factorization machines [1/2]

Matrix Factorization Training Data

	i_1	i_2	i_3
u_1	2	4	
u_2		1	
u_3	3		5

Factorization Machine Training Data

	u_1	u_2	u_3	i_1	i_2	i_3	a_1	a_2	y
x_1	1	0	0	1	0	0	2.0	0.0	2
x_2	1	0	0	0	1	0	1.5	0.5	4
x_3	0	1	0	0	1	0	0.0	1.0	1
x_4	0	0	1	1	0	0	0.3	0.7	3
x_5	0	0	1	0	0	1	3.2	1.7	5

Users

Items

Auxiliary Features

Observed Ratings

Factorization machines [2/2]

$$\hat{y}(\mathbf{x}) := w_0 + \sum_{i=1}^{n} w_i x_i + \sum_{i=1}^{n} \sum_{j=i+1}^{n} \langle \mathbf{v}_i, \mathbf{v}_j \rangle x_i x_j$$

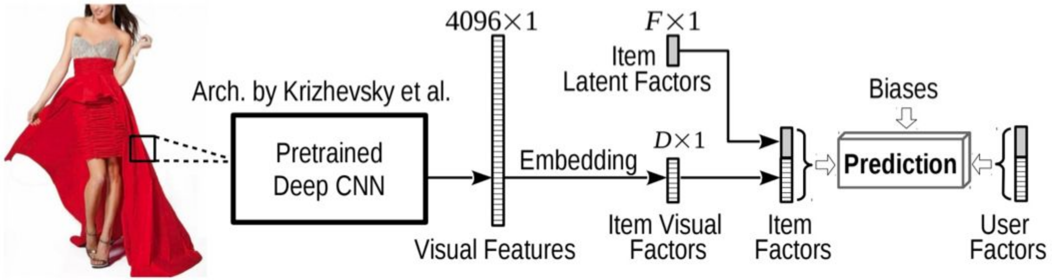
Coeficientes de
regresión de la
iesima variable

Factorización
Matricial
User - Item

Interacción
entre features
Xi y Xj

Imágenes en recomendaciones

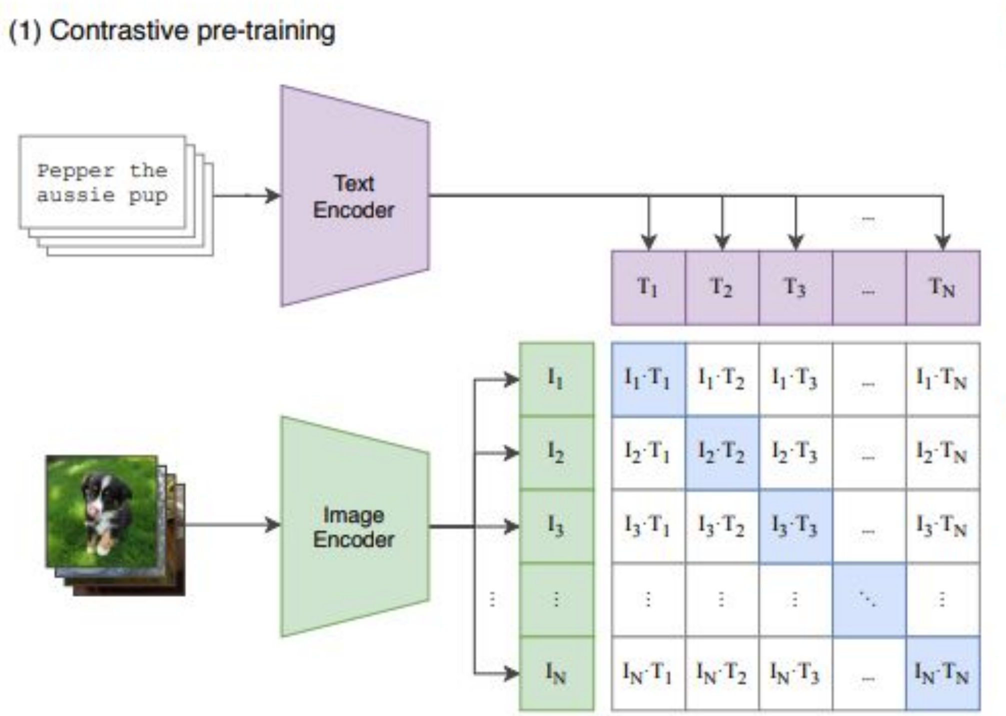
- VBPR: Visual Bayesian Personalized Ranking from Implicit Feedback [He and McAuley, 2016]
- Utiliza CNN para obtener características visuales
- Utilizan embeddings de usuarios e ítems
- Pérdida BPR (pairwise)



El Text Encoder y el Image Encoder se entrenan juntos para acercar imágenes y textos que corresponden entre sí.

El **contrastive pre-training** busca maximizar la similaridad coseno de la diagonal de la matriz N*N de los embeddings de imágenes y de textos.

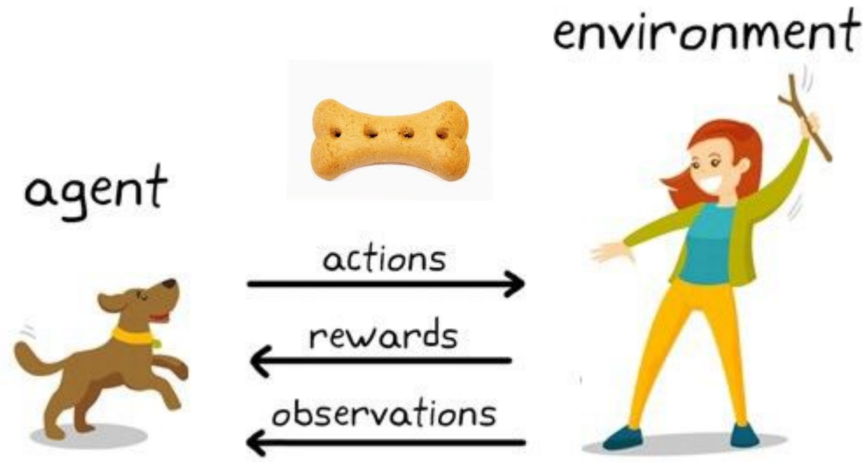
Se actualizan los pesos de ambos encoders.

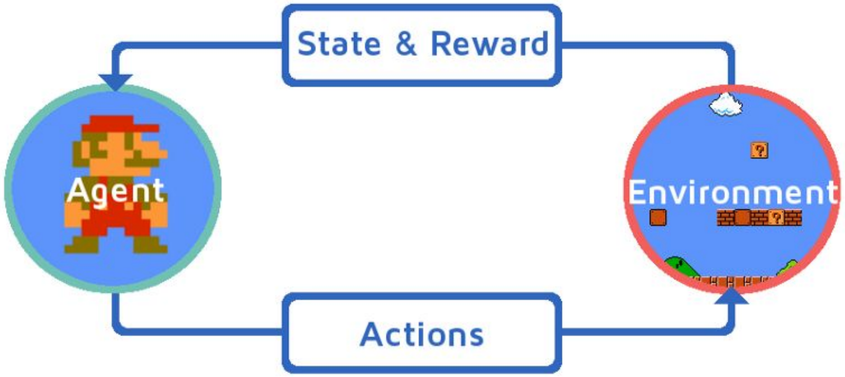



Product description
Product image
Related products
amazon
polo ralph lauren
Deliver to Singapore
Today's Deals
Your Amazon.com
Gift Cards
Help
Registry
Sell
Polo Ralph Lauren Mens Custom Slim Fit Mesh City Polo Shirt
295 ratings | 64 answered questions
Price: $54.75 - $134.99
Fit: As expected (79%)
Size:
Select
Size Chart
Color: Burgundy Houston
HOUSTON
100% Cotton
Polo Ralph Lauren
Custom Slim Fit
Features embroidered pony, city, and crest logos
Two riveted vent gromits under each arm
Mesh knit with tennis tails
Recommended from our brands
Amazon Essentials
$20.00
Amazon Essentials
$19.00
Havana Breeze
$29.99
Something for Everyone
$12.99 - $16.99
Customers who viewed this item also viewed
Page 1 of 5

APRENDIZAJE REFORZADO

APRENDIZAJE POR PRUEBA Y ERROR.
EXPLORAR ACCIONES NUEVAS APOSTANDO POR UNA RECOMPENSA A COSTA DE PERDER TIEMPO Y RECURSOS.





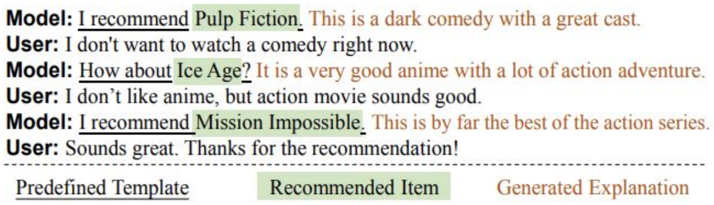
Sistemas recomendadores basados en bandits

Recomendación conversacional

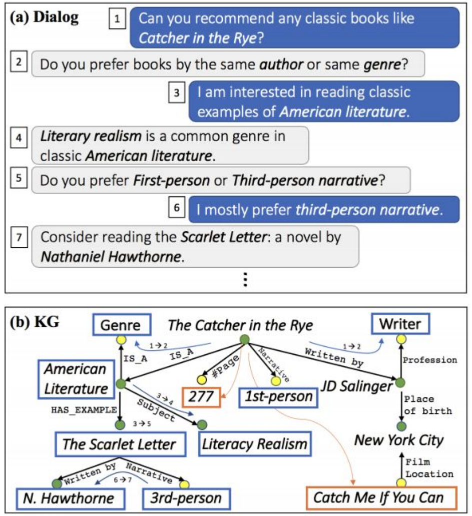
Sistema recomendador conversacional

¿Qué es un sistema conversacional?

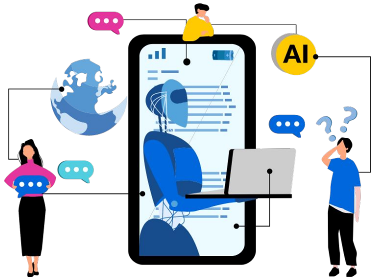
Un sistema recomendador que da **sugerencias personalizadas** a través de diálogos en lenguaje natural con un sistema.



CHEN ET AL, IJCAI 2020



OPENDIAL KG MOON ET AL, ACL 2019



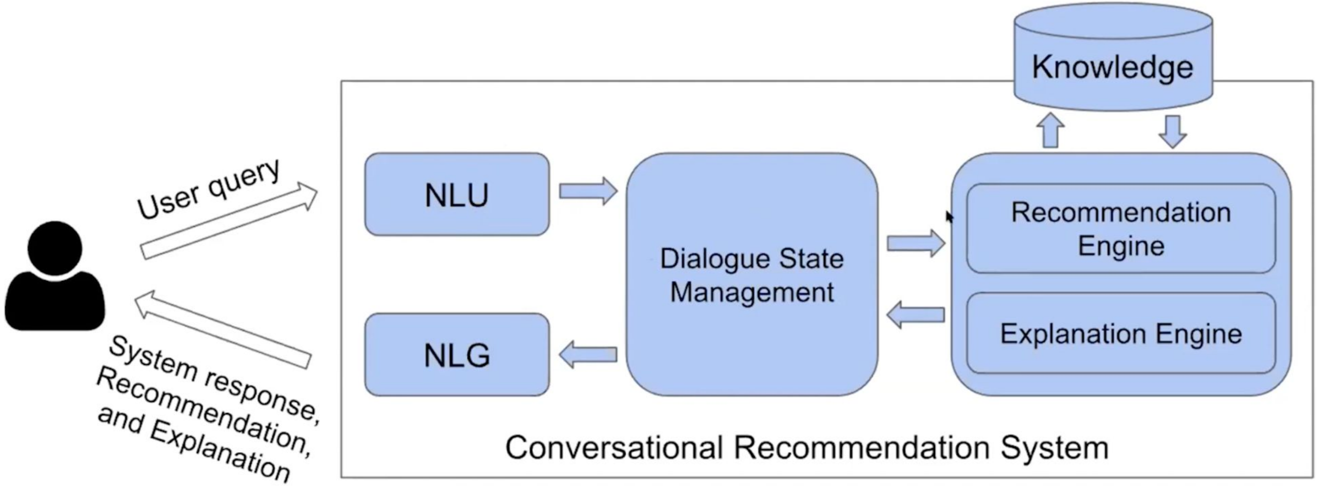
También puede funcionar en lenguaje **escrito** o **hablado**.

Sistema recomendador conversacional

una forma natural y privada de recibir recomendaciones:

- descubrir las necesidades del usuario mediante diálogos.
- dar orientación en una situación particular.

Sistema recomendador conversacional: arquitectura



Datos de entrenamiento

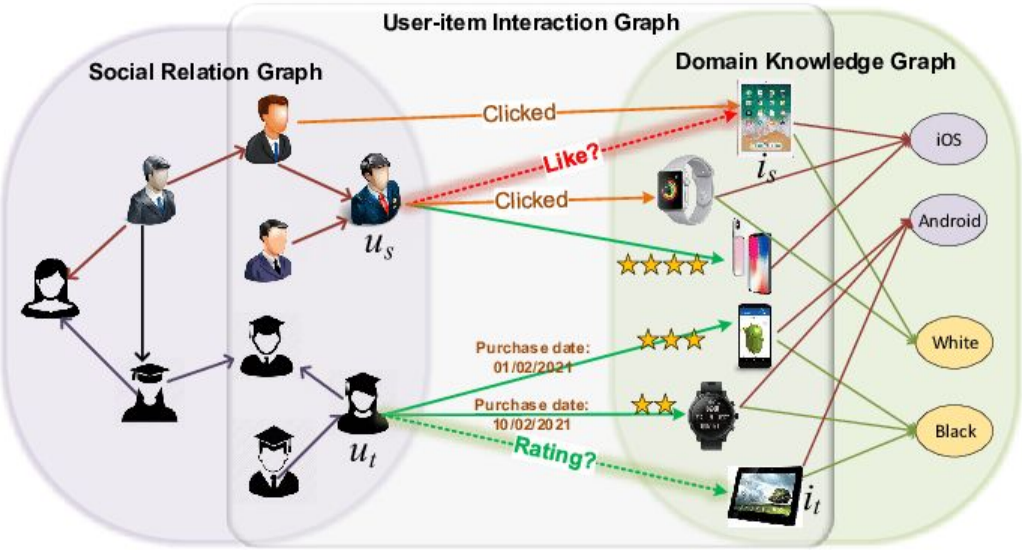
1. Convertir reviews de productos en una conversación.
2. Amazon mechanical turk - Diálogos entre dos usuarios.
3. Diálogos telefónicos transcritos.

…

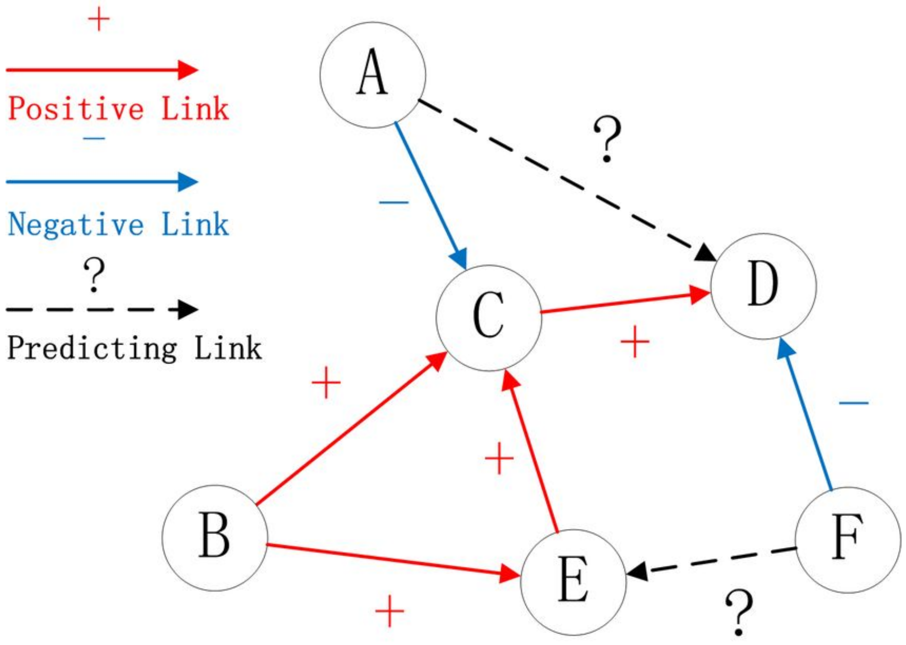
Dataset	Dialog Types	Domains	External Knowledge
ConvRec [10]	Rec	Restaurant	N
SAUR [11]	Rec	E-commerce	N
Cookie [24]	Rec	E-commerce	Y
ReDail [13]	Rec, Chitchat	Movie	N
OpendialKG [14]	Rec	Music, Sports	Y
KBRD [15]	Rec	Movie	Y
DuRecDial [21]	Rec, Chitchat, QA	Movie, Music, Restaurant, News, Weather	Y
MGConvRex [23]	Rec, Chitchat, QA	Restaurant	Y

Recomendación basada en grafos

Sistemas recomendadores basados en grafos



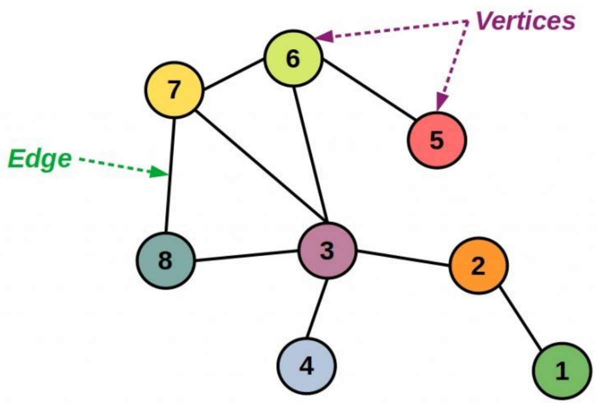
Una forma natural de representar interacciones en un sistema recomendador es con grafos.

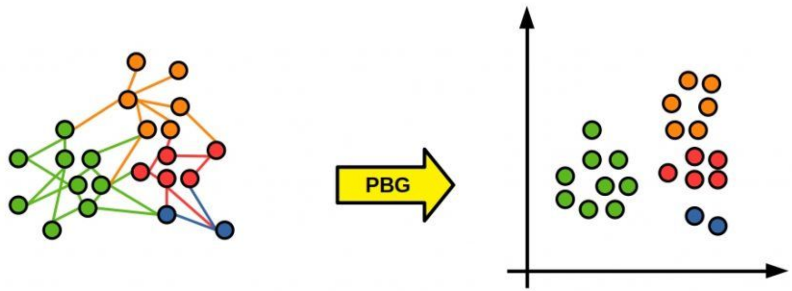


Aprender embeddings de usuarios e items (nodos dentro de un grafo)

A partir de las interacciones (aristas)

La tarea es predecir interacciones futuras entre nodos.





- Cada nodo es un vector de caracteristicas (embeddings)
- La tarea es actualizar estos vectores para forzarlos a predecir interacciones futuras con otros nodos.

Knowledge graph embedding learning

We have triplets **<s , r , d>** corresponding to E

$$< s, r^{(i)}, d >$$

Each element have vectors:

$$(\theta_s, \theta_r^{(i)}, \theta_d)$$

Function **f** relates **s, r** and **d**

$$f(\theta_s, \theta_r^{(i)}, \theta_d) = sim(g(\theta_s, \theta_r^{(i)}), g(\theta_d, \theta_r^{(i)}))$$

Objective: Maximize cases where the triple e belongs to E, and to minimize cases when the edge does not exists in the graph.

$$L = \sum_{e \in E} \sum_{e' \in S'_e} max(f(e) - f(e') + \lambda, 0)$$

Transformation function g encodes the relation between nodes and relations r (i) vectors. e’ is an edge that do not exist in the graph.

Knowledge graph embedding learning (relation functions g)

Model	Score function	#Parameters
SE [Bordes *et al.*, 2011]	$\Vert\mathbf{M_r^h h} - \mathbf{M_r^t t}\Vert_{\ell_{1/2}}$ $\mathbf{h}, \mathbf{t} \in \mathbb{R}^{k_e}, \mathbf{M_r^h}, \mathbf{M_r^t} \in \mathbb{R}^{k_e \times k_e}$	$2k_e n_e + 2k_e^2 n_r$
SME(linear) [Bordes *et al.*, 2012]	$-(\mathbf{M}_{h_1}\mathbf{e}_h + \mathbf{M}_{h_2}\mathbf{r} + \mathbf{b}_h)^\top(\mathbf{M}_{t_1}\mathbf{e}_t + \mathbf{M}_{r_2}\mathbf{r} + \mathbf{b}_t)$ $\mathbf{h}, \mathbf{t} \in \mathbb{R}^{k_e}, \mathbf{M_r^1}, \mathbf{M_r^2} \in \mathbb{R}^{k_e \times k_e}$	$2k_e n_e + 2k_e^2 n_r$
SME(bilinear) [Bordes *et al.*, 2012]	$-[(\mathbf{M}_{h_1}\mathbf{e}_h) \otimes (\mathbf{M}_{h_2}\mathbf{r}) + \mathbf{b}_h]^\top[(\mathbf{M}_{t_1}\mathbf{e}_t) \otimes (\mathbf{M}_{t_2}r) + \mathbf{b}_t]$ $\mathbf{h}, \mathbf{t} \in \mathbb{R}^{k_e}, \mathbf{M_r^1}, \mathbf{M_r^2} \in \mathbb{R}^{k_e \times k_e}$	$2k_e n_e + 2k_e^2 n_r$
TransE [Bordes *et al.*, 2013]	$\Vert\mathbf{h} + \mathbf{r} - \mathbf{t}\Vert_{\ell_{1/2}}$, $\mathbf{h}, \mathbf{t} \in \mathbb{R}^{k_e}, \mathbf{r} \in \mathbb{R}^{k_r}$	$k_e n_e + k_r n_r (*k_e = k_r)$
TransH [Wang *et al.*, 2014]	$\Vert\mathbf{h}(\mathbf{I} - \mathbf{w_r^\top w_t}) + \mathbf{r} - \mathbf{t}(\mathbf{I} - \mathbf{w_r^\top w_t})\Vert_{\ell_{1/2}}$, $\mathbf{h}, \mathbf{t} \in \mathbb{R}^k, \mathbf{w_r}, \mathbf{r} \in \mathbb{R}^k$	$k_e n_e + 2k_r n_r$
TransR [Lin *et al.*, 2015]	$\Vert\mathbf{hM_r} + \mathbf{r} - \mathbf{tM_r}\Vert_{\ell_{1/2}}$, $\mathbf{h}, \mathbf{t} \in \mathbb{R}^{k_e}, \mathbf{r} \in \mathbb{R}^{k_r}, \mathbf{M_r} \in \mathbb{R}^{k_e \times k_r}$	$k_e n_e + (k_r + k_r^2) n_r$
TransD [Ji *et al.*, 2015]	$\Vert(\mathbf{r}_p\mathbf{h}_p^\top + \mathbf{I}^{k_r \times k_e})\mathbf{h} + \mathbf{r} - (\mathbf{r}_p\mathbf{t}_p^\top + \mathbf{I}^{k_r \times k_e})\mathbf{t}\Vert_{\ell_{1/2}}$, $\mathbf{h}, \mathbf{t}, \mathbf{h_p}, \mathbf{t_p} \in \mathbb{R}^{k_e}, \mathbf{r}, \mathbf{r_p} \in \mathbb{R}^{k_r}$	$4k_e n_e + 2k_r n_r$
TranSparse(*separate*) *[Ji* et al.*, 2016]*	$\Vert\mathbf{M_r^1}(\theta_r^1)\mathbf{h} + \mathbf{r} - \mathbf{M_r^2}(\theta_r^2)\mathbf{t}\Vert_{\ell_{1/2}}$ $\mathbf{h}, \mathbf{t} \in \mathbb{R}^{k_e}, r \in \mathbb{R}^{k_r}, \mathbf{M_r^1}(\theta_r^1), \mathbf{M_r^2}(\theta_r^2) \in \mathbb{R}^{*k_r \times k_e}$	$2k_e n_e + 2(1-\theta)(k_e + 1)k_r n_r$ $0 \ll \theta_r^1, \theta_r^2 \leq 1$
GTrans-DW [Tan *et al.*, 2018]	$\Vert 1/\sigma \odot [(\alpha\mathbf{h}_e + \beta\mathbf{r}_a\mathbf{h}_a^\top\mathbf{h}_e) + \mathbf{r}_e - (\alpha\mathbf{t}_e + \beta\mathbf{r}_a\mathbf{t}_a^\top\mathbf{t}_e)]\Vert_{\ell_{1/2}}$ $\mathbf{h}_a, \mathbf{h}_e, \mathbf{t}_a, \mathbf{t}_e \in \mathbb{R}^{k_e}, \mathbf{r}_a, \mathbf{r}_e \in \mathbb{R}^{k_r}, \sigma, \alpha, \beta \in \mathbb{R}$	$k_e n_e + 3k_r n_r (*k_e = k_r)$
GTrans-SW [Tan *et al.*, 2018]	$\Vert 1/\sigma \odot [(1 - \alpha_{h,r})\mathbf{h}_e + \alpha_{h,r}\mathbf{r}_a\mathbf{h}_a^\top\mathbf{h}_e + \mathbf{r}_e - (1 - \alpha_{t,r})\mathbf{t}_e$ $-\alpha_{t,r}\mathbf{r}_a\mathbf{t}_a^\top\mathbf{t}_e]\Vert_{\ell_{1/2}}\ \mathbf{h}_a, \mathbf{h}_e, \mathbf{t}_a, \mathbf{t}_e \in \mathbb{R}^{k_e}, \mathbf{r}_a, \mathbf{r}_e \in \mathbb{R}^{k_r}, \sigma, \alpha, \beta \in \mathbb{R}$	$k_e n_e + 3k_r n_r (*k_e = k_r)$
TransMS	$\Vert -\tanh(\mathbf{t} \otimes \mathbf{r}) \otimes \mathbf{h} + \mathbf{r} - \tanh(\mathbf{h} \otimes \mathbf{r}) \otimes \mathbf{t} + \alpha \cdot g(\mathbf{h} \otimes \mathbf{t})\Vert_{\ell_{1/2}}$, $\mathbf{h}, \mathbf{t} \in \mathbb{R}^{k_e}, \mathbf{r} \in \mathbb{R}^{k_r}, \alpha = \mathbf{r}_{k_r+1} \in \mathbb{R}^1$	$k_e n_e + (k_r + 1) n_r (*k_e = k_r)$

Table 2: Statistics of datasets used in experiments. We mainly compare the models' score functions and their numbers of parameters. n_e and n_r represent the number of entities and relations in knowledge graph respectively. k_e and k_r represent the dimension of entity and relation in the low-dimensional space, $\mathbf{h}, \mathbf{t} \in \mathbb{R}^{k_e}$, $\mathbf{r} \in \mathbb{R}^{k_r}$. θ_r^1 and θ_r^1 denote the sparse degrees of transfer matrices in TranSparse. α represents one additional dimension for each relation vectors in our model TransMS. For TransE, GTrans and TransMS, ($*k_e = k_r$) means that k_e is equal to k_r.

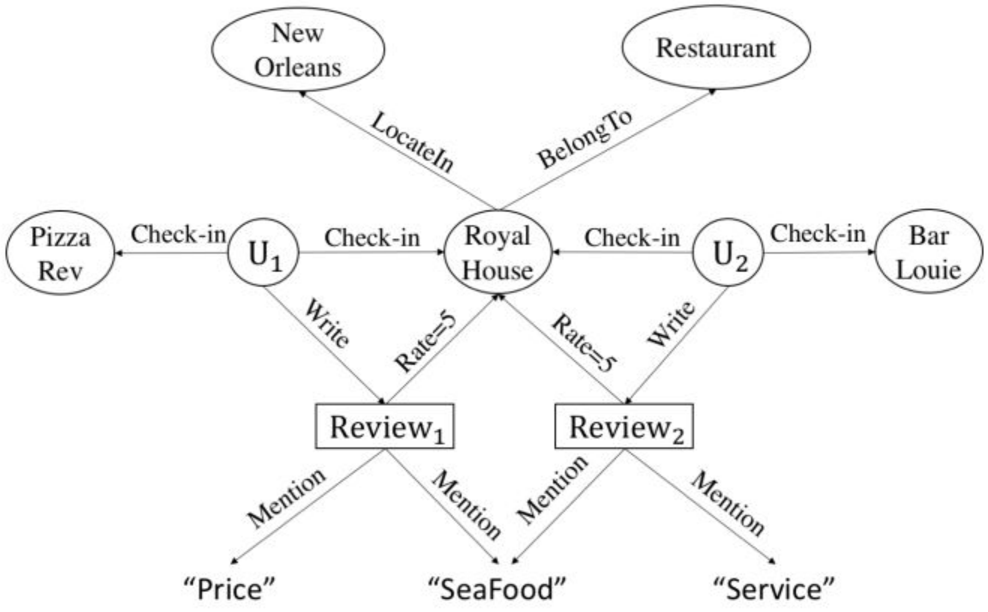


Figure 1: Example of HIN, which is built based on the web page for Royal House on Yelp.

Gracias!