



Convolutional Matrix Factorization for Document Context-Aware Recommendation

*Donghyun Kim, Chanyoung Park, Jinhoh Oh, Sungyoung Lee,
Hwanjo Yu*

Proceedings of the 10th ACM Conference on Recommender
Systems

September, 2016



Introducción



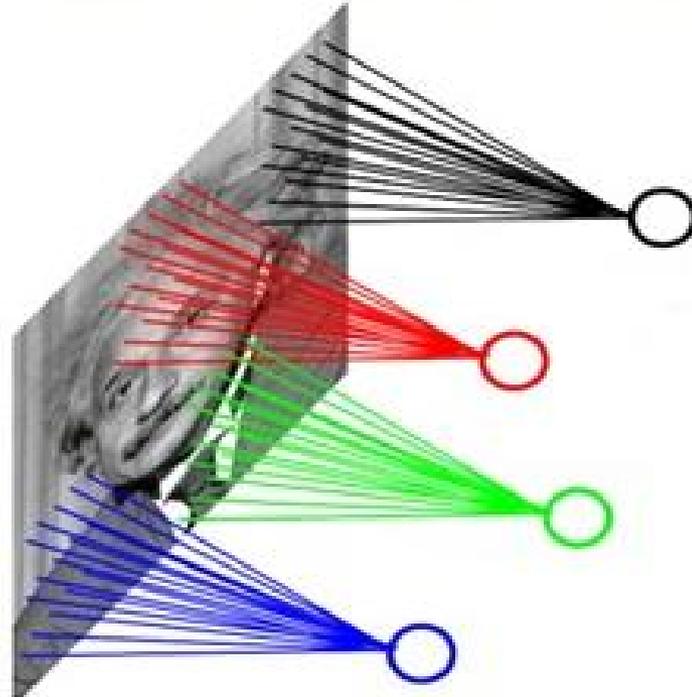
Introducción

- Problema
 - Calidad de recomendaciones es mermada por la baja densidad de datos
 - Creación de sistemas híbridos que incorporen información contextual (contenido: descripción de ítems)
 - Modelo BoW posee limitaciones inherentes al modelo
- Propuesta
 - modelo de recomendación basado en contexto utilizando ConvMF
 - ConvMF = CNN + PMF



Introducción

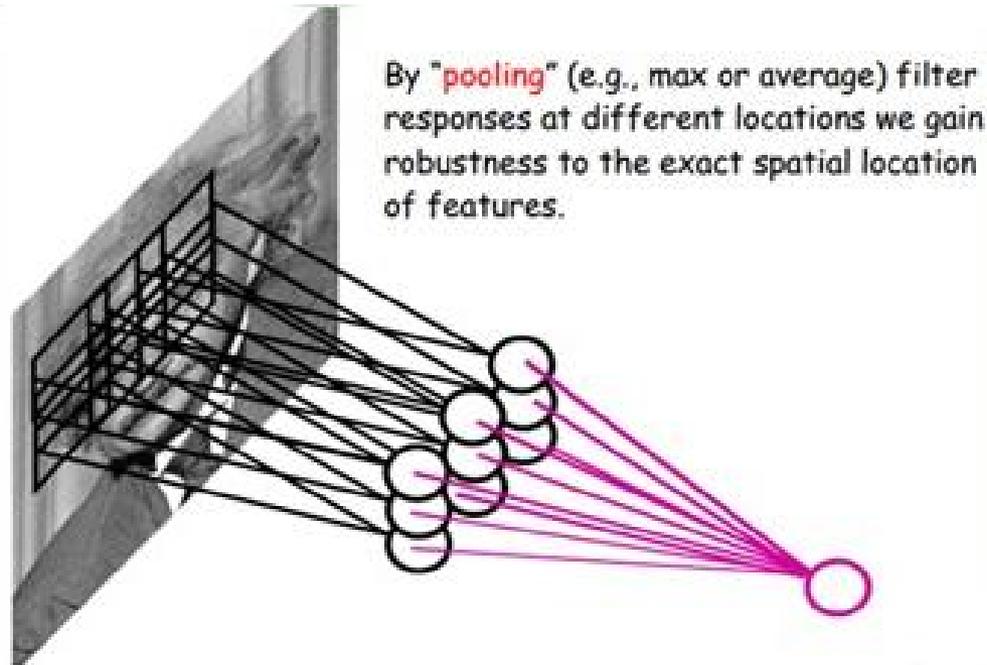
Redes Neuronales Convolucionales





Redes Neuronales Convolucionales

- Pooling





Factorización Matricial Convolutiva

ConvMF



Modelo probabilístico de ConvMF

- Probabilistic Matrix Factorization (PMF)
- Sea $R \in \mathbb{R}^{N \times M}$ matriz de ratings
- Con N usuarios y M ítems
- Se desea encontrar $U \in \mathbb{R}^{k \times N}$ y $V \in \mathbb{R}^{k \times M}$ tal que

$$R \approx U^T V$$



Factorización Matricial Probabilística

- Modelo de regresión lineal con ruido Gaussiano

$$r_{ui} = \mathbf{u}_i^T \mathbf{v}_j + \varepsilon$$

- Distribución condicional sobre los ratings observados

$$p(R|U, V, \sigma^2) = \prod_i^N \prod_j^M N(r_{ij} | \mathbf{u}_i^T \mathbf{v}_j, \sigma^2)^{I_{ij}}$$



Factorización Matricial Convolutiva

Factorización Matricial Probabilística

- Se asigna modelo gaussiano de media cero a distribuciones de probabilidad a priori

$$p(U|\sigma_U^2) = \prod_{i=1}^N \mathcal{N}(U_i|0, \sigma_U^2 \mathbf{I}), \quad p(V|\sigma_V^2) = \prod_{j=1}^M \mathcal{N}(V_j|0, \sigma_V^2 \mathbf{I})$$



Factorización Matricial Probabilística

- Se asigna modelo gaussiano de media cero a distribuciones de probabilidad a priori

$$p(U|\sigma_U^2) = \prod_{i=1}^N \mathcal{N}(U_i|0, \sigma_U^2 \mathbf{I}), \quad p(V|\sigma_V^2) = \prod_{j=1}^M \mathcal{N}(V_j|0, \sigma_V^2 \mathbf{I})$$



Modelo latente de ítems

- Es generado a partir de variables
 1. Pesos de la CNN
 2. Representación X_j del documento asociado ítem j
 3. Ruido gaussiano: ε

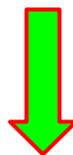


Modelo latente de ítems

$$v_j = \text{cnn}(W, X_j) + \epsilon_j$$

$$\epsilon_j \sim N(0, \sigma_V^2 I)$$

$$p(W|\sigma_W^2) = \prod_k N(w_k|0, \sigma_W^2)$$

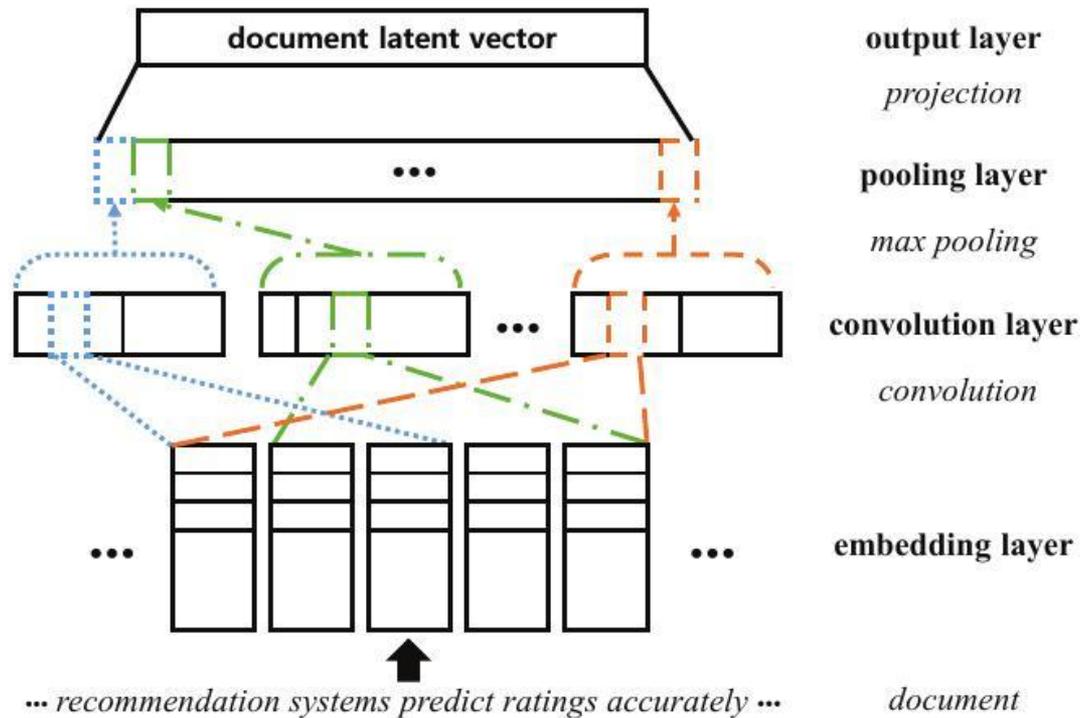


$$p(U|\sigma_U^2) = \prod_{i=1}^N \mathcal{N}(U_i|0, \sigma_U^2 \mathbf{I}), \quad p(V|W, X, \sigma_V^2) = \prod_j^M N(v_j|\text{cnn}(W, X_j), \sigma_V^2 I)$$



Factorización Matricial Convolutiva

Arquitectura de CNN





Factorización Matricial Convolutional

Embedding Layer

- Entrada: documento crudo
- Salida: secuencia de embeddings de palabra
- Inicialización: aleatoria o modelo tipo GloVe

$$D = \begin{bmatrix} \cdots & w_{i-1} & w_i & w_{i+1} & \cdots \end{bmatrix}$$



Factorización Matricial Convolutiva

Convolution Layer

- Extracción de características contextuales $c_i^j \in \mathbb{R}$
- Ponderados por pesos compartidos $W_c^j \in \mathbb{R}^{p \times w_s}$
- Función de activación: ReLU

$$c_i^j = f(W_c^j * D_{(:, i:(i+w_s-1))} + b_c^j)$$



Factorización Matricial Convolutiva

Convolution Layer

- Vector de características contextuales $c^j \in \mathbb{R}^{l-ws+1}$ para el documento W_c^j

$$c^j = [c_1^j, c_2^j, \dots, c_i^j, \dots, c_{l-ws+1}^j]$$

- Se utilizan múltiples pesos compartidos para extraer más de un tipo de características contextuales: $j = 1, 2, 3 \dots n_c$



Factorización Matricial Convolutiva

Pooling Layer

- Extracción de características representativas
- Construcción de vector de tamaño fijo
- Max-pooling

$$d_f = [\max(c^1), \max(c^2), \dots, \max(c^j), \dots, \max(c^{n_c})]$$



Factorización Matricial Convolutiva

Output Layer

- Proyección de vector d_f a espacio k -dimensional
- Generación de vector latente de documento

$$s = \tanh(W_{f_2} \{ \tanh(W_{f_1} d_f + b_{f_1}) \} + b_{f_2})$$

- Con $W_{f_1} \in \mathbb{R}^{f \times n_c}$, $W_{f_2} \in \mathbb{R}^{k \times f}$



Experimentos



Experimentos

Dataset

- Amazon y MovieLens 1m y 10m de ratings (más IMDb)
 - Largo máximo de documentos: 300
 - Eliminación de stop-words
 - Cálculo de tf-idf y eliminación de frecuencias mayores a 0.5
 - Tamaño de vocabulario: 8000 palabras
-



Competidores

- Probabilistic Matrix Factorization (PMF)
 - Collaborative Topic Regression (CTR)
 - Collaborative Deep Learning (CDL)
 - Convolutional Matrix Factorization (ConvMF)
 - Convolutional Matrix Factorization con embedding pre-entrenado (ConvMF+)
-



Detalles de implementación

- Vectores latentes de palabras de dimensión 200
- Múltiples tamaños de ventana en capa convolucional: 3, 4, y 5
- 100 pesos compartidos por tamaño de ventana
- Se utiliza *dropout* en vez de regularizador



Experimentos

Test RMSE

Model	Dataset		
	ML-1m	ML-10m	AIV
PMF	0.8971	0.8311	1.4118
CTR	0.8969	0.8275	1.5496
CDL	0.8879	0.8186	1.3594
ConvMF	0.8531	0.7958	1.1337
ConvMF+	0.8549	0.7930	1.1279
Improve	3.92%	2.79%	16.60%

Dataset	# users	# items	# ratings	density
ML-1m	6,040	3,544	993,482	4.641%
ML-10m	69,878	10,073	9,945,875	1.413%
AIV	29,757	15,149	135,188	0.030%



Experimentos

Análisis de densidad de dataset

Model	Ratio of training set to the entire dataset (density)						
	20% (0.93%)	30% (1.39%)	40% (1.86%)	50% (2.32%)	60% (2.78%)	70% (3.25%)	80% (3.71%)
PMF	1.0168	0.9711	0.9497	0.9354	0.9197	0.9083	0.8971
CTR	1.0124	0.9685	0.9481	0.9337	0.9194	0.9089	0.8969
CDL	1.0044	0.9639	0.9377	0.9211	0.9068	0.8970	0.8879
ConvMF	0.9745	0.9330	0.9063	0.8897	0.8726	0.8676	0.8531
Improve	2.98%	3.20%	3.36%	3.41%	3.77%	3.27%	3.92%



Referencias

- [1] D. Kim, C. Park, J. Oh, S. Lee, H. Yu, Convolutional Matrix Factorization for Document Context-Aware Recommendation, in Proceedings of the 10th ACM Conference on Recommender Systems, 2016.
- [2] R. Salakhutdinov and A. Mnih. Probabilistic matrix factorization. In Advances in Neural Information Processing Systems, volume 20, 2008.
- [3] C. Rasmussen and C. Williams, Gaussian Processes for Machine Learning (Adaptive Computation and Machine Learning), 2005.