

Sistemas de Recomendación en Netflix Grand Prize

Miguel Miní

Facultad de Ciencias, Ciencias de la
computación

Universidad Nacional de Ingeniería

mminih@uni.pe

I. INTRODUCCIÓN

Los algoritmos de recomendación y en general los sistemas de recomendación son un área de mucha investigación en las últimas décadas, en especial desde la aparición de los primeros algoritmos de filtro colaborativo a mediados de la década de los 90 [1]. Hoy en día Tienen una gran popularidad en los *e-commerce web sites* tales como *Amazon*, *Spotify* y *Netflix*, donde se usa un *input* del cliente para generar una lista de items recomendados. Muchas aplicaciones solo usan los *items* que el cliente compra y que califica explícitamente para representar sus intereses, pero ellos puedes además usar otros atributos, incluyendo vistas de *items*, datos demográficos, temas de interés, y artistas favoritos [2].

II. STATE OF THE ART

Los sistemas de recomendación aparecen por primera vez en arduos trabajos sobre ciencia cognitiva [3], teoría de la aproximación [4], recuperación de la información [5] y predicción de teorías [6], los sistemas de recomendación surgen a mediados de la década de los 90 cuando los investigadores inician enfocándose sobre los problemas de recomendación que explícitamente confían sobre la estructura del *rating*. En la formulación más común, el problema de recomendación es reducido para el problema de estimar el *rating* para los *items* que no han sido visto por un usuario.

Formalmente los sistemas de recomendación pueden ser formulados como sigue: Sea C el conjunto de todos los usuarios y sea S el conjunto de todos los items posibles que pueden ser recomendados, tal como películas, CDs y libros. Sea u sea una función de utilidad tal que mide lo

util de un item s es para un usuario, formalmente:

$$\forall c \in C, s_c = \arg \max_{s \in S} u(c, s)$$

Donde la recomendación es s_c , la función u suele basarse en recomendación entre usuarios similares o *items* similares.

El problema central de los sistemas de recomendación residen en el hecho de que no tenemos información de la utilidad de todos los *items* para todos los usuarios, pero si para un subconjuntos de estos por lo cual tenemos que interpolar esos resultados.

Los nuevos *rating* de *items* que no han sido obtenidos pueden ser estimados en varias maneras usando metodos desde machine learning, deep learning, teoría de la aproximación y varias heurísticas, entre las 3 principales, tenemos: [7].

- * *Content-based recommendations*: El usuario es recomendado con *items* desde *items* que eligió en el pasado;
- * *Collaborative recommendations*: El usuario es recomendado *items* que otros usuarios con gustos similares han preferido en el pasado.
- * *Hybrid approaches*: Estos metodos combinan *Collaborative* y *Content-based recommendations*

Fig. 1. collaborative-based filtering

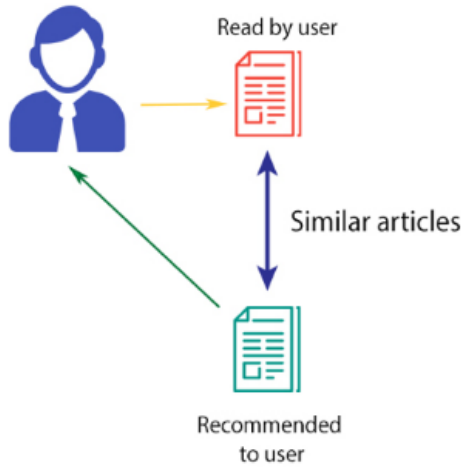


Fig. 2. content-based filtering

A. Netflix Prize

En setiembre 2009, Netflix en su concurso Grand Netflix Prize premio al equipo que mejoraba su actual predicción en un 10% con data de películas desde su plataforma recolectada desde sus usuarios. El algoritmo con mejores resultados fue *collaborative filtering*, basado en vecindad (*k-nearest neighbors*) el cual fue mejorado por el equipo ganador "BellKor's Pragmatic Chaos"[11] [12] [14] [13] [15] [16] llevándose \$1.0M. [8]

Rank	Team Name	Best Test Score	% Improvement	Best Submit Time
Grand Prize - RMSE = 0.8567 - Winning Team: BellKor's Pragmatic Chaos				
1	BellKor's Pragmatic Chaos	0.8567	10.06	2009-07-26 18:18:28
2	The Ensemble	0.8567	10.06	2009-07-26 18:38:22
3	Grand Prize Team	0.8582	9.90	2009-07-10 21:24:40

Fig. 3. laderboard Netflix Grand Prize [17]

B. Redes Neuronales

En 2017, 6 investigadores hicieron un analisis extensivo con redes neuronales para *collaborative filtering* [9], donde se explica como el avance de algoritmos de deep learning suelen ser usados como algoritmos auxiliares en los sistemas de recomendaciones, tales como reconocimiento de sonidos, y relaciones entre los usuarios y los elementos. Además proponen un framework para sistema de recomendación de filtro colaborativo basado en redes neuronales (NCF) y observaciones de mejoras significativas en el desempeño de este

framework en comparación con métodos más tradicionales.

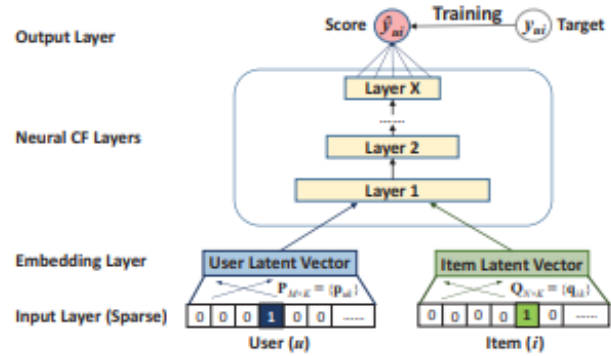


Fig. 4. Neuronal collaborative filtering framework

C. Mi Proyecto

En los siguientes avances comparare algoritmos de sistemas de recomendacion, en especial *collaborative filtering* basado en redes neuronales y en vecindad como aquellos que se usaron en el Netflix Grand Prize, para ello dispondré de la data que liberó Netflix en su concurso Netflix Grand Prize, obtenida de la página "www.kaggle.com". Todo el proyecto se subirá a "www.github.com"[10] en mi repositorio[18].

REFERENCES

- [1] Gediminas Adomavicius, Member, IEEE, and Alexander Tuzhilin, Member, IEEE. Toward the Next Generation of Recommender Systems: A Survey of the State-of-the-Art and Possible Extensions, 2005
- [2] Greg Linden, Brent Smith, and Jeremy York. Amazon.com Recommendations Item-to-Item Collaborative Filtering. Amazon. 2003.
- [3] Elaine Rich. The University of Texas at Austin. User Modeling via Stereotypes. 1979.
- [4] M.J.D. Powell, Approximation Theory and Methods. Cambridge Univ. Press, 1981
- [5] G. Salton, Automatic Text Processing. Addison-Wesley, 1989.
- [6] J. Scott Armstrong. Principles of Forecasting A Handbook for Researchers and Practitioner, 2001.
- [7] Balabanovi, M., & Shoham, Y. . Fab: Content-Based, Collaborative Recommendation. 1997.
- [8] Netflix Inc. Netflix Prize. "https://www.netflixprize.com/index.html". 2009. accessed 16-11-2018.
- [9] Xiangnan He, Lizi Liao, Hanwang Zhang, Liqiang Nie, Xia Hu, Tat-Seng Chua. Neural Collaborative Filtering. 2017.
- [10] Kaggle "https://www.kaggle.com/netflix-inc/netflix-prize-data". 2018. accessed 16-11-2018.
- [11] Robert M. Bell and Yehuda Koren. Scalable collaborative filtering with jointly derived neighborhood interpolation-weights. Seventh IEEE International Conference on Data Mining, 2007.

- [12] Michael Jahrer Andreas Toscher and Robert M. Bell. The bigchaos solution to the netflix grand prize.commendoresearch & consulting, 2009
- [13] Yehuda Koren. The bellkor solution to the netflix grand prize. 2009.
- [14] Martin Pottie and Martin Chabbert. The pragmatic theory solution to the netflix grand prize. 2009
- [15] Yehuda Koren Robert M. Bell and Chris Volinsky. The bellkor 2008 solution to the netflix prize. 2008.
- [16] Andreas Toscher and Michael Jahrer. The bigchaos solution to the netflix prize 2008.commendo research & consulting, 2008.
- [17] Netflix Inc. Netflix Prize. "<https://www.netflixprize.com/leaderboard.html>". 2009. accessed 16-11-2018.
- [18] Github Inc. miguelAlessandro. "<https://github.com/miguelAlessandro/Curso-CM072>". 2018. accesed 16-11-2018