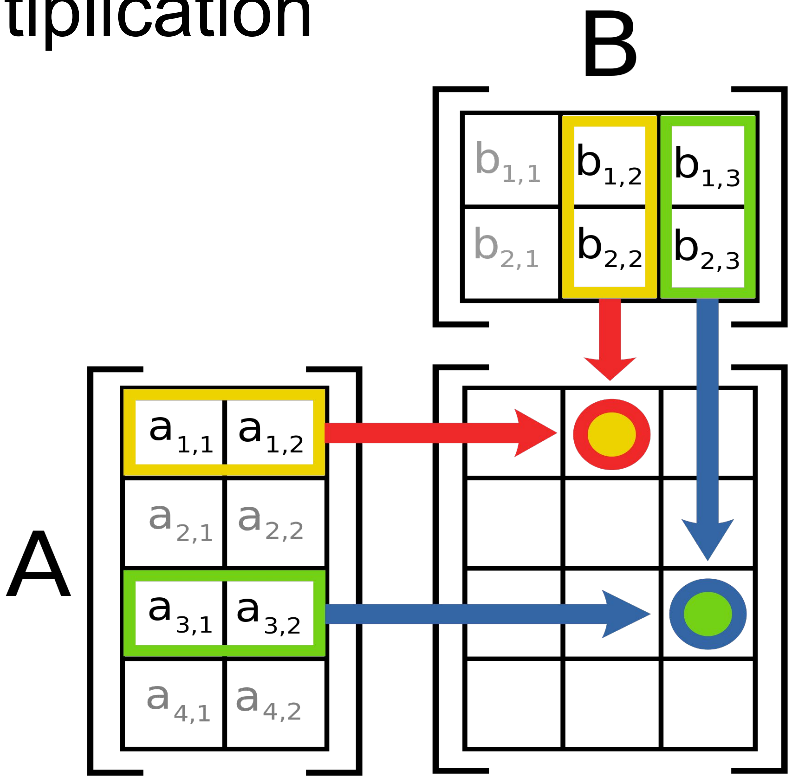
Matrix Factorisation for Recommender Systems

Contents

- Refresher about matrices
- SVD
- NMF
- ALS
- Recommender Metrics

Matrix Multiplication

Matrix Multiplication



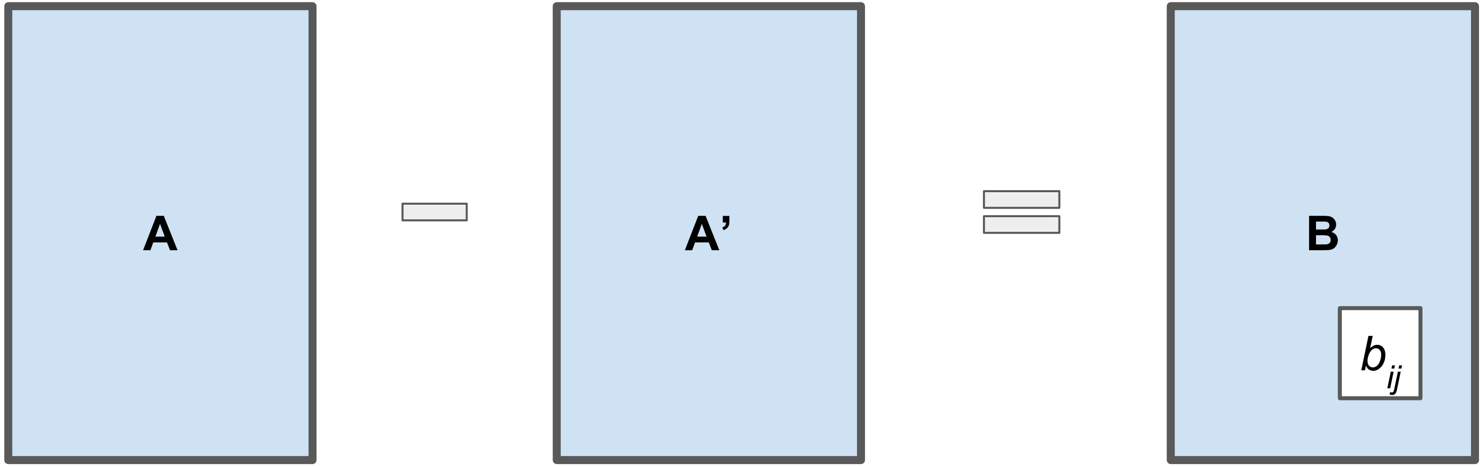
Matrix Multiplication

$I_n \rightarrow$ identity matrix size n

$\forall\ A_{m \times n},\ \exists\ I_n, I_m\text{:}\ \ AI_n = I_mA = A$

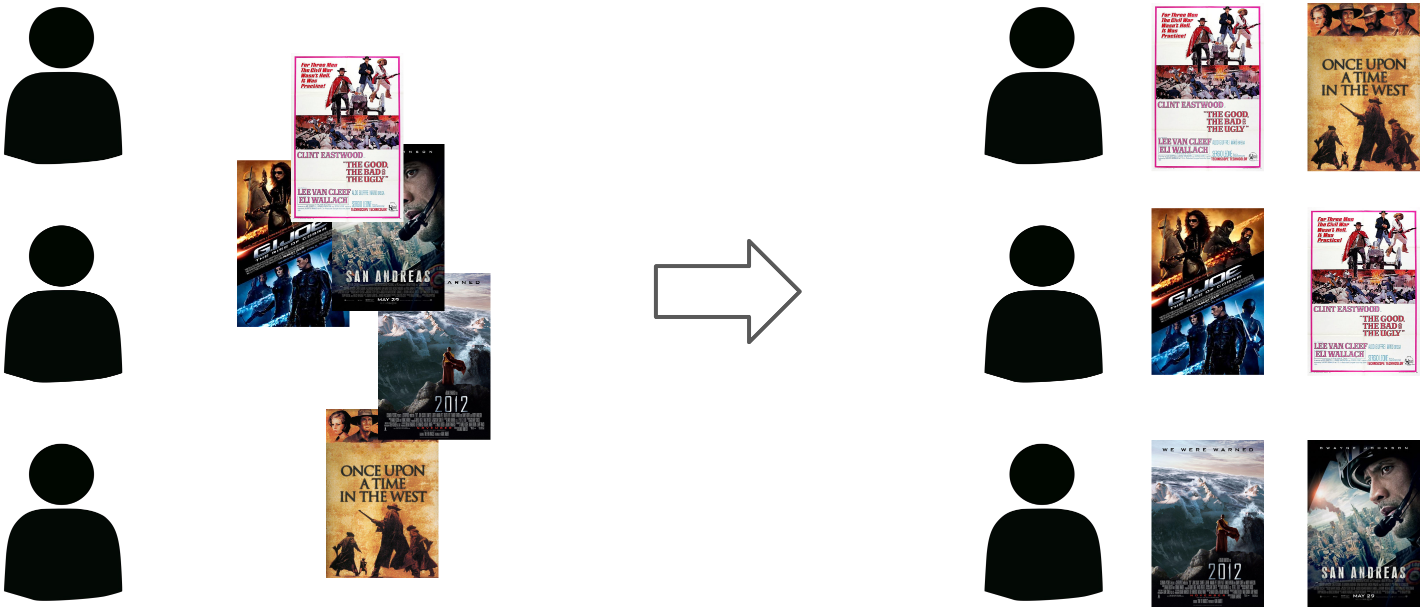
$$I_1 = \begin{bmatrix} 1 \end{bmatrix},\ I_2 = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix},\ I_3 = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix},\ \cdots,\ I_n = \begin{bmatrix} 1 & 0 & 0 & \cdots & 0 \\ 0 & 1 & 0 & \cdots & 0 \\ 0 & 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & 1 \end{bmatrix}$$

Least Squares

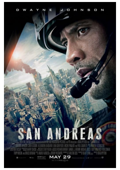


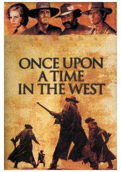
$$\|B\|_F = \sqrt{\sum_{i=0}^{m} \sum_{j=0}^{n} |b_{ij}|^2}$$

The Problem


For Three Men The Civil War Wasn't Hell. It Was Practice!
CLINT EASTWOOD
"THE GOOD, THE BAD & THE UGLY"
LEE VAN CLEEF
ELI WALLACH
SAN ANDREAS
MAY 29
2012
ONCE UPON A TIME IN THE WEST
WE WERE WARNED
DWAYNE JOHNSON

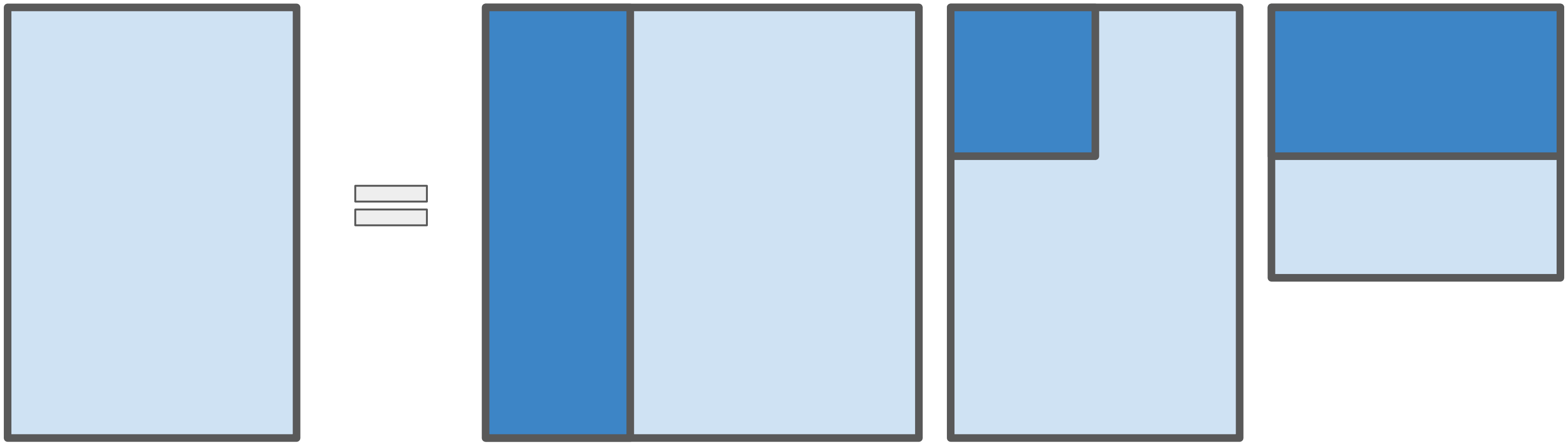
The Problem





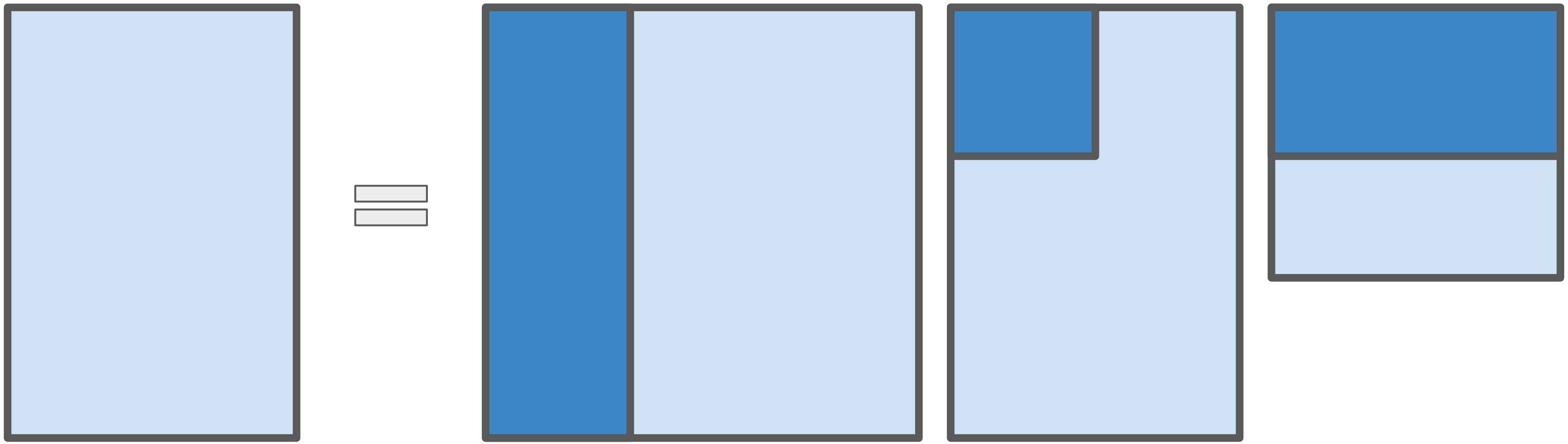
★★★★★	N/A	N/A	★☆☆☆☆	N/A
N/A	N/A	★★★☆☆	N/A	N/A
N/A	★★★☆☆	N/A	N/A	★★★★☆

SVD



A	**U**	**Σ**	**V^T**
$(m \times n)$	$(m \times m)$ or $(m \times k)$	$(m \times n)$ or $(k \times k)$	$(n \times n)$ or $(k \times n)$

SVD



https://github.com/mehese/svd_als_presentation/blob/master/SVD.ipynb

~~Everything should be made as simple as possible, but not simpler.~~

~~Albert Einstein~~

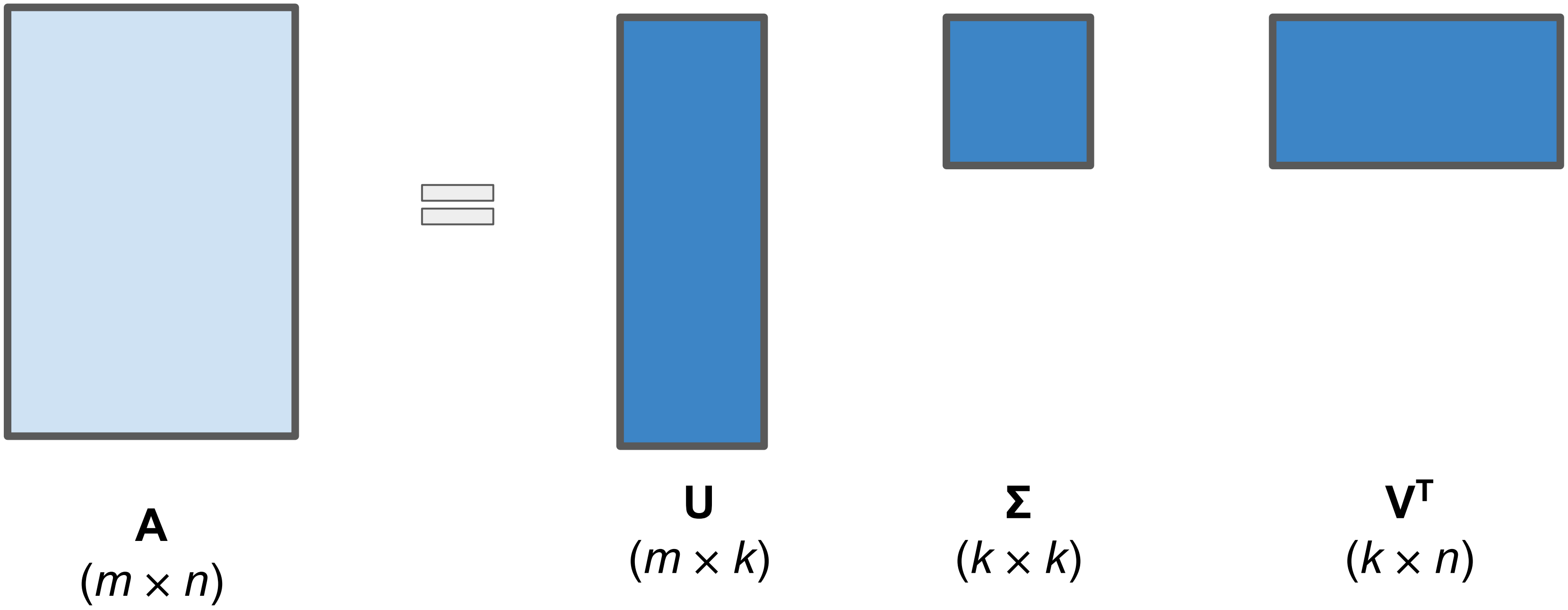
It can scarcely be denied that the supreme goal of all theory is to make the irreducible basic elements as simple and as few as possible without having to surrender the adequate representation of a single datum of experience.

Albert Einstein

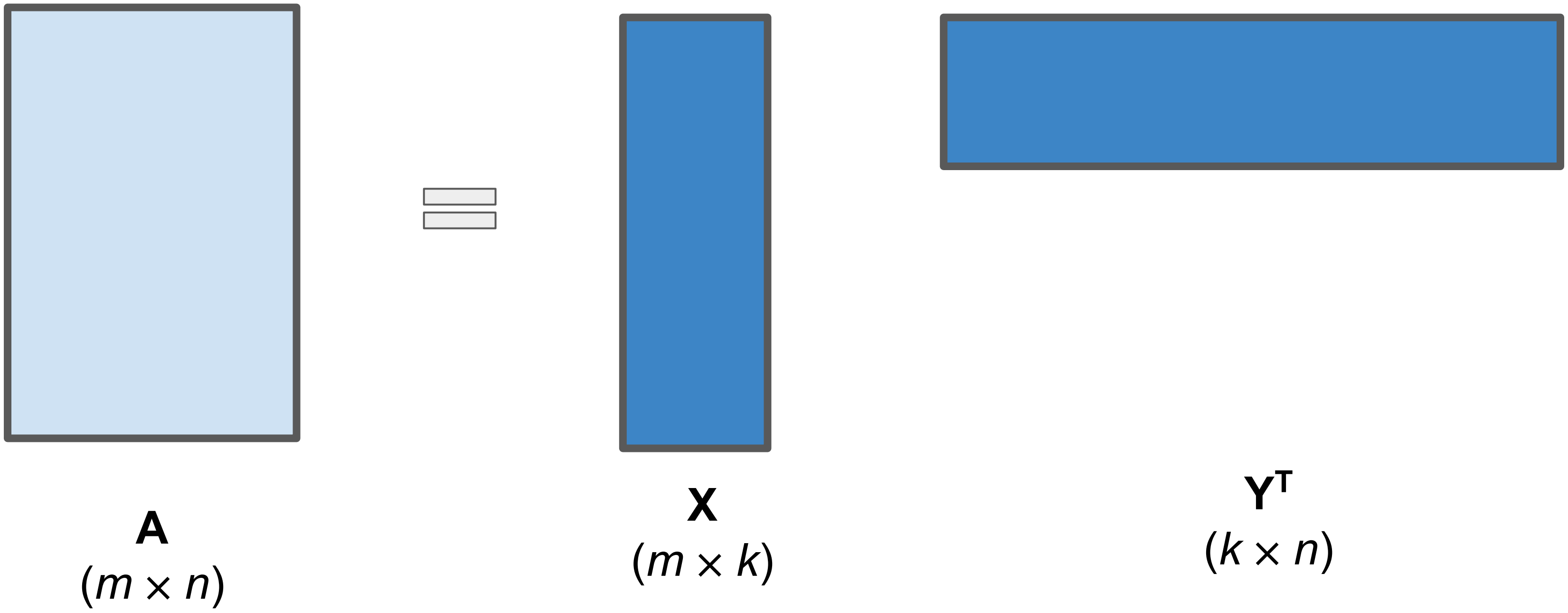
"On the Method of Theoretical Physics" The Herbert Spencer Lecture, delivered at Oxford (10 June 1933); also published in *Philosophy of Science*, Vol. 1, No. 2 (April 1934), pp. 163-169., p. 165

(but I got it from Wikipedia and never double-checked…)

NMF



NMF



ALS

Collaborative Filtering for Implicit Feedback Datasets

Yifan Hu
AT&T Labs – Research
Florham Park, NJ 07932

Yehuda Koren*
Yahoo! Research
Haifa 31905, Israel

Chris Volinsky
AT&T Labs – Research
Florham Park, NJ 07932

Abstract

A common task of recommender systems is to improve customer experience through personalized recommendations based on prior implicit feedback. *These systems passively track different sorts of user behavior, such as purchase history, watching habits and browsing activity, in order to model user preferences. Unlike the much more extensively researched* explicit feedback, *we do not have any direct input from the users regarding their preferences. In particular, we lack substantial evidence on which products consumer dislike. In this work we identify unique properties of implicit feedback datasets. We propose treating the data as indication of positive and negative preference associated with vastly varying confidence levels. This leads to a factor model which is especially tailored for implicit feedback recommenders. We also suggest a scalable optimization procedure, which scales linearly with the data size. The algorithm is used successfully within a recommender system for television shows. It compares favorably with well tuned implementations of other known methods. In addition, we offer a novel way to give explanations to recommendations given by this factor model.*

1 Introduction

As e-commerce is growing in popularity, an important challenge is helping customers sort through a large variety of offered products to easily find the ones they will enjoy the most. One of the tools that address this challenge is recommender systems, which are attracting a lot of attention recently [1, 4, 12]. These systems provide users with per-

tent based approach creates a profile for each user or product to characterize its nature. As an example, a movie profile could include attributes regarding its genre, the participating actors, its box office popularity, etc. User profiles might include demographic information or answers to a suitable questionnaire. The resulting profiles allow programs to associate users with matching products. However, content based strategies require gathering external information that might not be available or easy to collect.

An alternative strategy, our focus in this work, relies only on past user behavior without requiring the creation of explicit profiles. This approach is known as *Collaborative Filtering* (CF), a term coined by the developers of the first recommender system - Tapestry [8]. CF analyzes relationships between users and interdependencies among products, in order to identify new user-item associations. For example, some CF systems identify pairs of items that tend to be rated similarly or like-minded users with similar history of rating or purchasing to deduce unknown relationships between users and items. The only required information is the past behavior of users, which might be their previous transactions or the way they rate products. A major appeal of CF is that it is domain free, yet it can address aspects of the data that are often elusive and very difficult to profile using content based techniques. While generally being more accurate than content based techniques, CF suffers from the *cold start* problem, due to its inability to address products new to the system, for which content based approaches would be adequate.

Recommender systems rely on different types of input. Most convenient is the high quality *explicit feedback*, which includes explicit input by users regarding their interest in products. For example, Netflix collects star ratings

Y Hu et al, 2008

COVER FEATURE

MATRIX FACTORIZATION TECHNIQUES FOR RECOMMENDER SYSTEMS

Yehuda Koren, *Yahoo Research*
Robert Bell and Chris Volinsky, *AT&T Labs—Research*

As the Netflix Prize competition has demonstrated, matrix factorization models are superior to classic nearest-neighbor techniques for producing product recommendations, allowing the incorporation of additional information such as implicit feedback, temporal effects, and confidence levels.

Modern consumers are inundated with choices. Electronic retailers and content providers offer a huge selection of prod-

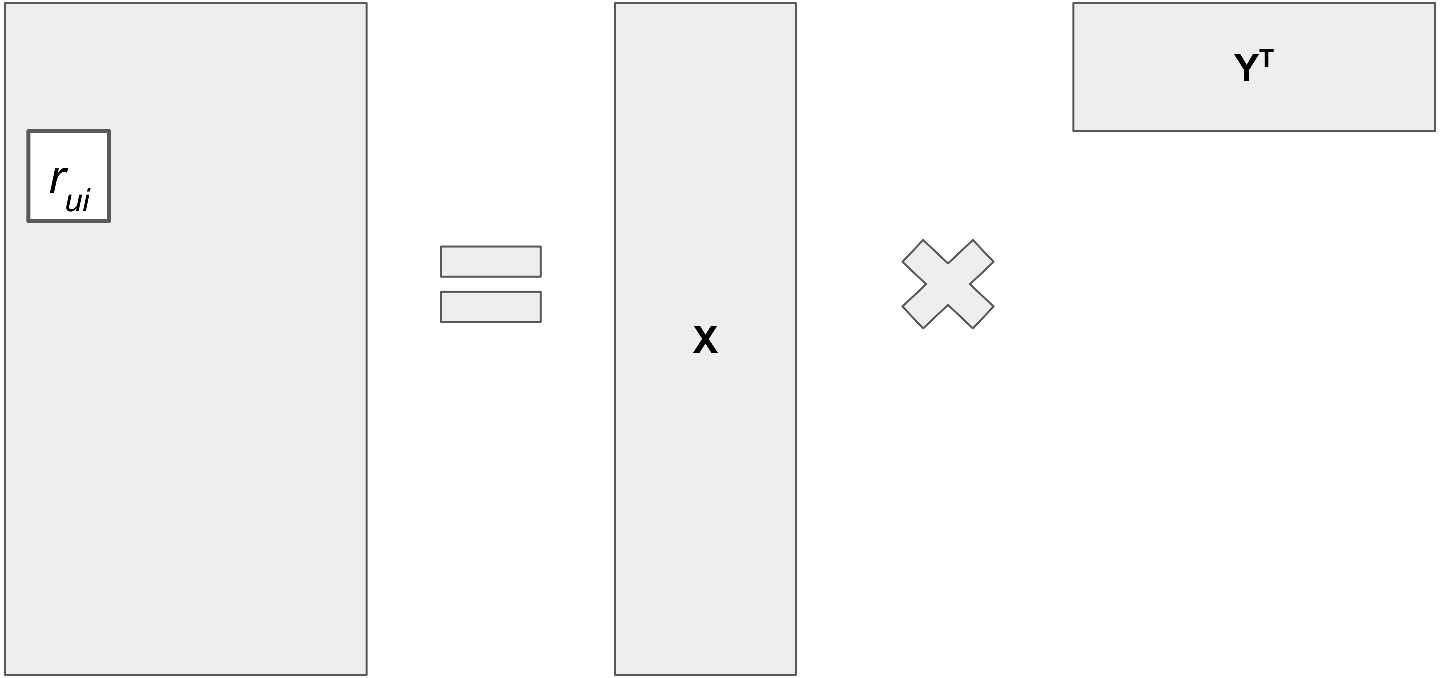
Such systems are particularly useful for entertainment products such as movies, music, and TV shows. Many customers will view the same movie, and each customer is likely to view numerous different movies. Customers have proven willing to indicate their level of satisfaction with particular movies, so a huge volume of data is available about which movies appeal to which customers. Companies can analyze this data to recommend movies to particular customers.

RECOMMENDER SYSTEM STRATEGIES

Broadly speaking, recommender systems are based on one of two strategies. The *content filtering* approach

Y Koren et al, 2009

ALS



ALS

$$\min_{x^*,y^*} \sum_{\text{known } r_{ui}} \left(r_{ui} - x_u^T y_i\right)^2 - \lambda \left(\|x_u\|^2 + \|y_i\|^2\right)$$

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

$$p_{ui} = \begin{cases} 1 & , r_{ui} > 0 \\ 0 & , r_{ui} = 0 \end{cases}$$

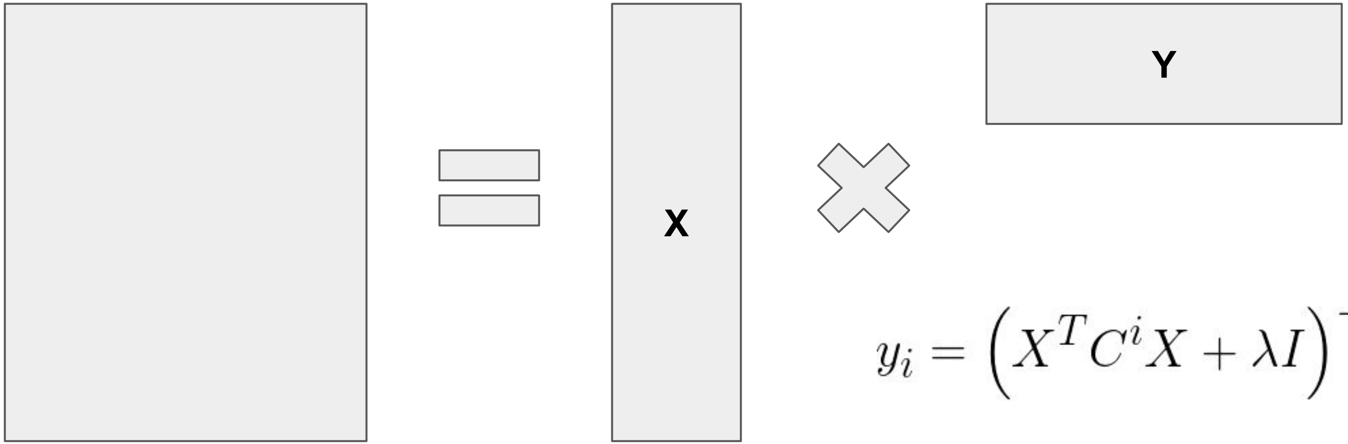
$$\min_{x^*y^*} \sum_{u,i} c_{ui} \left(p_{ui} - x_u^T y_i\right)^2 - \lambda \left(\sum_u \|x_u\|^2 + \sum_i \|y_i\|^2\right)$$

ALS

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

$$x_u = \left(Y^T C^u Y + \lambda I\right)^{-1} Y^T C^u p(u) \qquad y_i = \left(X^T C^i X + \lambda I\right)^{-1} X^T C^i p(i)$$

ALS



$$y_i = \left(X^T C^i X + \lambda I\right)^{-1} X^T C^i p(i)$$

$$x_u = \left(Y^T C^u Y + \lambda I\right)^{-1} Y^T C^u p(u)$$

https://github.com/mehese/svd_als_presentation/blob/master/ALS.ipynb

ALS

How to get even more speedup

$$x_u = \left(Y^T C^u Y + \lambda I\right)^{-1} Y^T C^u p(u)$$

$$Y^T C^u Y = Y^T Y + Y^T (C^u - I)$$

Small number of non-zero elements

ALS

But what about implicit feedback?

r_{ui} = ★★☆☆☆ ?!?

$$\hat{r}_{ui} = \mu + b_i + b_u + q_i^T [p_u + |N(u)|^{-0.5} \sum_{i \in N(u)} x_i + \sum_{a \in A(u)} y_a]$$

Metrics

General

- Precision
- Root Mean Square Error

Rank aware

- Mean Average Precision
- Normalised Discounted Cumulative Gain

Others

- Reach
- Serendipity

Resources

- https://github.com/fastai/numerical-linear-algebra/blob/master/README.md
- http://ieeexplore.ieee.org/document/4781121/
- http://ieeexplore.ieee.org/abstract/document/5197422/
- https://www.cs.cmu.edu/~venkatg/teaching/CStheory-infoage/book-chapter-4.pdf
- https://www.youtube.com/watch?v=mBcLRGuAFUk
- https://fenix.tecnico.ulisboa.pt/downloadFile/3779576344458/singular-value-decomposition-fast-track-tutorial.pdf

Thank you

Eric Mehes

https://www.linkedin.com/in/eric-mehes/

@meHessian