Advanced Machine Learning

Modern Recommender Systems

Rodrigo Da Silva Alves

March 06, 2025

Dimensionality Reduction

- Technique to reduce features in a dataset while retaining most relevant information
- Different inputs (e.g., music, photo, text) have unique characteristics requiring specific machine learning approaches
- Traditional methods like PCA may fail, especially those with non-linear relationships
- Autoencoder: Unsupervised neural network used for compressed data representation
  - Effective for dimensionality reduction and handling complex inputs.

RECAP: Autoencoder

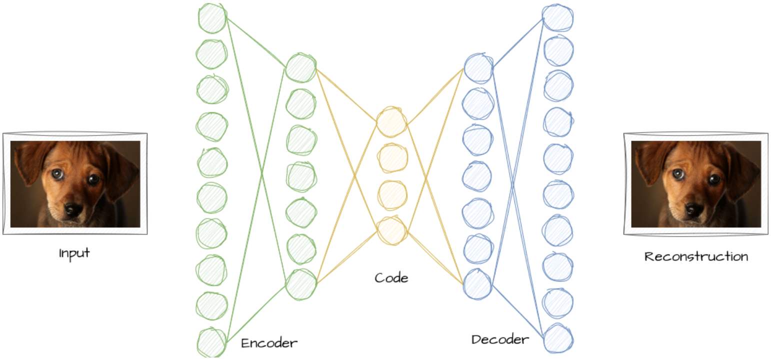
- An autoencoder is a type of feed-forward neural network
- It is designed to reconstruct its input x_i ad output x_i
- To prevent trivial solutions, the network includes a bottleneck (or code) layer
  - Significantly smaller dimension than the input

RECAP: Autoencoder

- An autoencoder is also composed by a encoder/decoder
- The encoder and the decoder have normally similar structure
- More formally: let $\mathcal{E}()$ be a encoder and $\mathcal{D}()$ be a decoder. Our optmization problem can be described as:

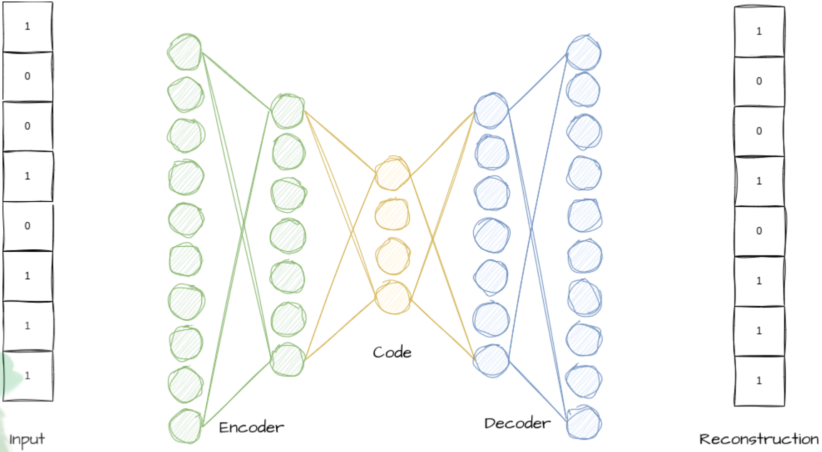
$$\text{min}_{\mathcal{E},\mathcal{D}} \sum_{i} ||x_i - \mathcal{D}(\mathcal{E}(x_i))||$$

Autoencoder



How can we use autoencoders to predict implicit feedback?

Autoencoders for CF



Autoencoders for CF

- Autoencoders are frequently used for collaborative filtering.
- They are very accurate in predicting rankings.
- They can also be used to find clusters with the code.
- Empirical results show that the best architecture is often not very deep.
- What would be the shallowest autoencoder for Collaborative Filtering?

EASE

- EASE is the shallowest auto-encoder as possible
- It aims to solve the following problem

$$\text{min}_B ||X - XB||^2 + \lambda ||B||^2 \text{ s.t } \text{diag}(B) = 0$$

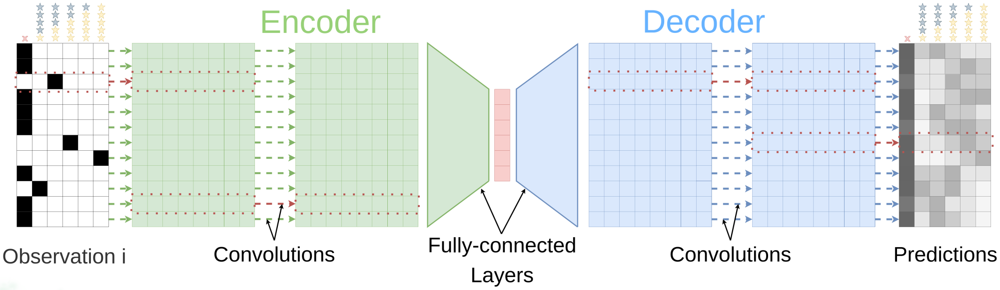
- Why do we need the constraint $\text{diag}(B) = 0$?
- EASE has closed form solution! See here
- Is this a good method?

EASE Results

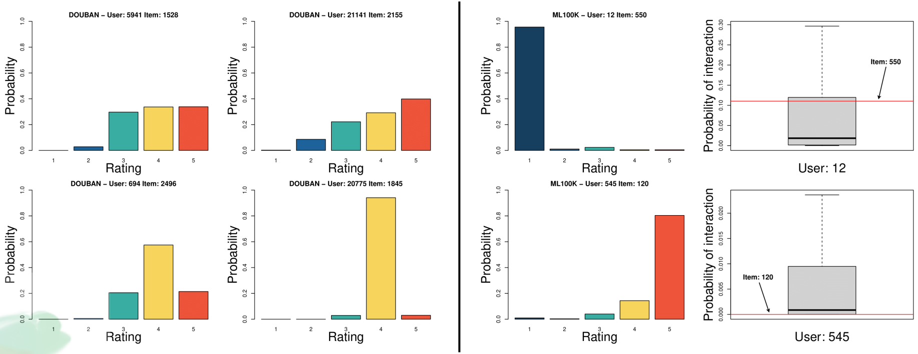
Table 1: Ranking accuracy (with standard errors of about 0.002, 0.001, and 0.001 on the *ML-20M*, *Netflix*, and *MSD* data, respectively), following the experimental set-up in [13].

(a) *ML-20M*	Recall@20	Recall@50	NDCG@100
popularity	0.162	0.235	0.191
EASE^R	0.391	0.521	0.420
EASE$^R \geq 0$	0.373	0.499	0.402
results reproduced from [13]:			
SLIM	0.370	0.495	0.401
WMF	0.360	0.498	0.386
CDAE	0.391	0.523	0.418
MULT-VAE ^{PR}	0.395	0.537	0.426
MULT-DAE	0.387	0.524	0.419
(b) *Netflix*			
popularity	0.116	0.175	0.159
EASE^R	0.362	0.445	0.393
EASE$^R \geq 0$	0.345	0.424	0.373
results reproduced from [13]:			
SLIM	0.347	0.428	0.379
WMF	0.316	0.404	0.351
CDAE	0.343	0.428	0.376
MULT-VAE ^{PR}	0.351	0.444	0.386
MULT-DAE	0.344	0.438	0.380
(c) *MSD*			
popularity	0.043	0.068	0.058
EASE^R	0.333	0.428	0.389
EASE$^R \geq 0$	0.324	0.418	0.379
results reproduced from [13]:			
SLIM	— did not finish in [13] —		
WMF	0.211	0.312	0.257
CDAE	0.188	0.283	0.237
MULT-VAE ^{PR}	0.266	0.364	0.316
MULT-DAE	0.266	0.363	0.313

Mixing Implicit and Explicit



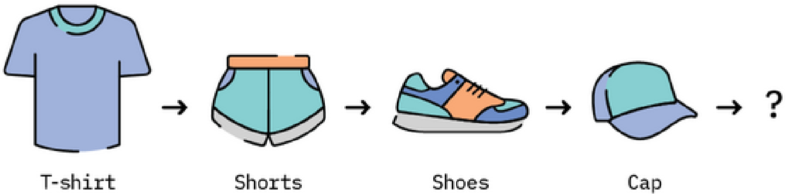
Mixing Implicit and Explicit



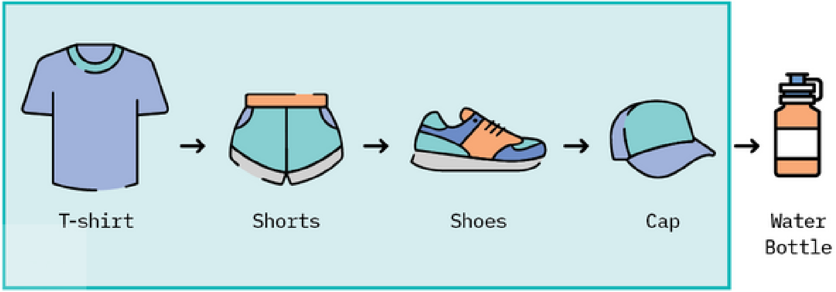
Sequential Recommendation

- Sequential recommendation is the task of predicting the next item that a user will interact
- There is extensive sequential recommendation algorithms
  - Markov chains
  - Recurrent neural networks (RNNs)
  - Long short-term memory (LSTM) networks
  - Embedding-base Neural Networks
- The models should learn patterns in a user's behavior over time

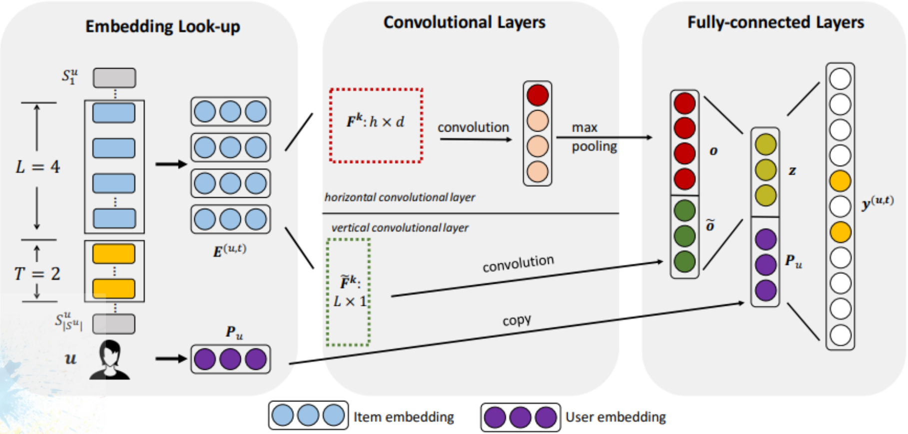
Sequential Recommendation



Sequential Recommendation



Sequential Recommendation

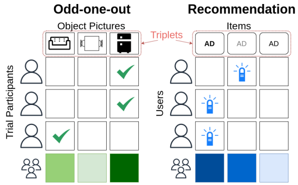


Triplets problem



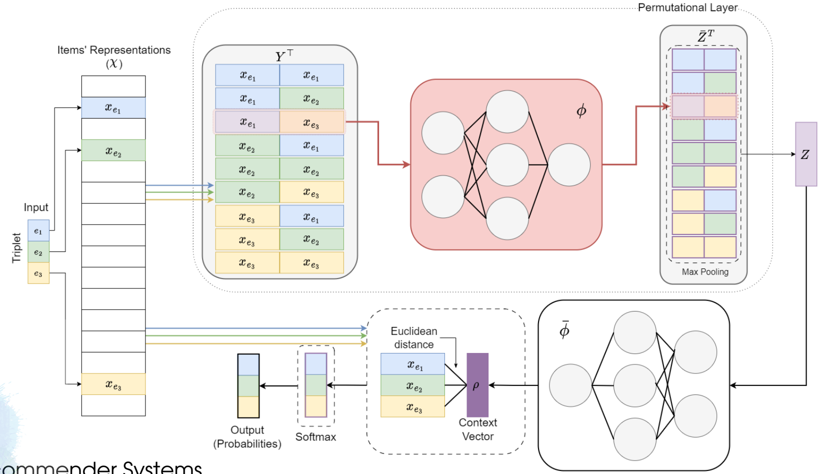



Triplets problem to Recommenders



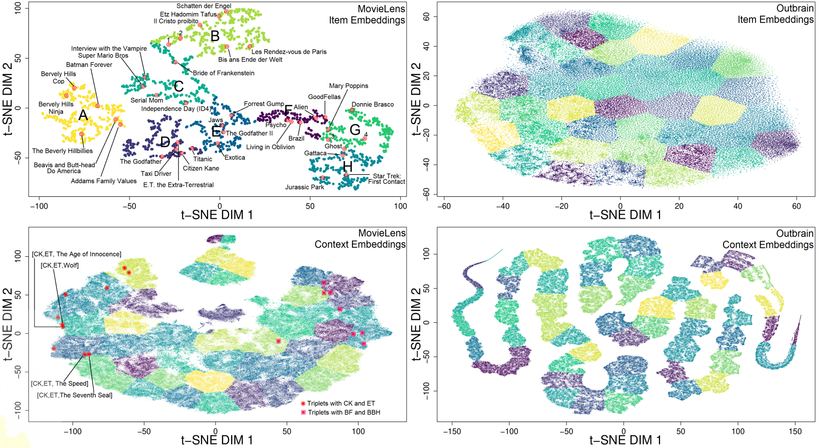
- The items we show to user can influence their decision
- Based on neuroscience
- Sometimes the position we show does not matter significantly
- Context embedding: summarizes the context of the recommendation
- Provide not just accurate recommendation but also interpretability

Care Model



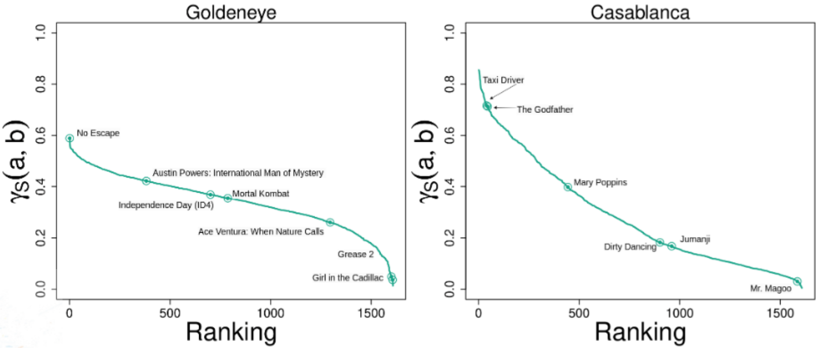
Permutational Layer
Items' Representations
($\mathcal{X}$)
$Y^\top$
ϕ
$\bar{Z}^T$
Max Pooling
Z
Input
Triplet
e_1
e_2
e_3
x_{e_1}
x_{e_2}
x_{e_3}
Euclidean distance
ρ
Context Vector
$\bar{\phi}$
Output (Probabilities)
Softmax

Care Model



MovieLens
Item Embeddings
Schatten der Engel
Etz Hadomim Tafus
Il Cristo proibito
Interview with the Vampire
Super Mario Bros
Batman Forever
Bervely Hills
Cop
Bervely Hills
Ninja
Serial Mom
Independence Day (ID4)
Les Rendez-vous de Paris
Bis ans Ende der Welt
Bride of Frankenstein
Forrest Gump
Alien
Mary Poppins
GoodFellas
Donnie Brasco
Jaws
Psycho
The Godfather II
Brazil
Living in Oblivion
Ghost
Gattaca
Exotica
Titanic
Citizen Kane
The Godfather
Taxi Driver
E.T. the Extra-Terrestrial
The Beverly Hillbillies
Beavis and Butt-head
Do America
Addams Family Values
Jurassic Park
Star Trek:
First Contact
A
B
C
D
E
F
G
H
t–SNE DIM 1
t–SNE DIM 2
Outbrain
Item Embeddings
MovieLens
Context Embeddings
[CK,ET, The Age of Innocence]
[CK,ET,Wolf]
[CK,ET, The Speed]
[CK,ET,The Seventh Seal]
Triplets with CK and ET
Triplets with BF and BBH
Outbrain
Context Embeddings

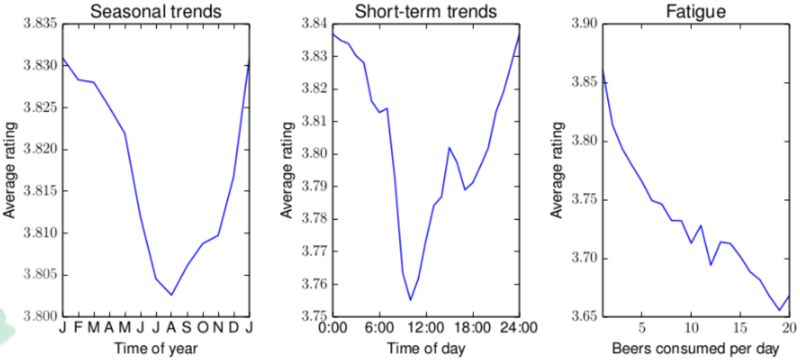
Care Model


Goldeneye
γs(a, b)
Ranking
No Escape
Austin Powers: International Man of Mystery
Mortal Kombat
Independence Day (ID4)
Ace Ventura: When Nature Calls
Grease 2
Girl in the Cadillac
0
500
1000
1500
0.0
0.2
0.4
0.6
0.8
1.0
Casablanca
γs(a, b)
Ranking
Taxi Driver
The Godfather
Mary Poppins
Dirty Dancing
Jumanji
Mr. Magoo
0
500
1000
1500
0.0
0.2
0.4
0.6
0.8
1.0

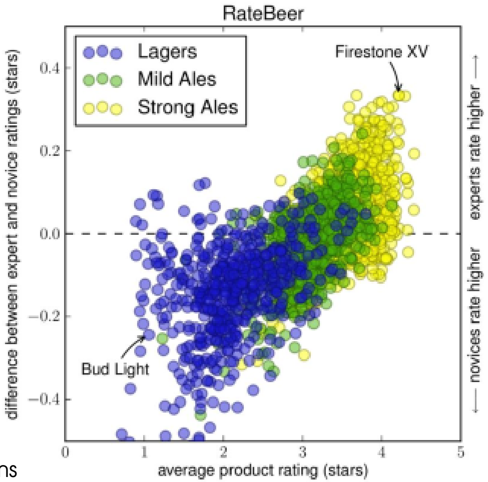
The Time Dimension in Recommendations

- Do you like the same things, morning and evenings?
- For example, the playlist recommendations on Spotify should change based on the time of day and day of the week.
  - Rarely do people have the same mood on Monday morning as they do on Friday evening.
- Taste and preferences change over time, so recommendations should adapt accordingly.
- The envoroment of RS is dynamic

Pivo Recommendation



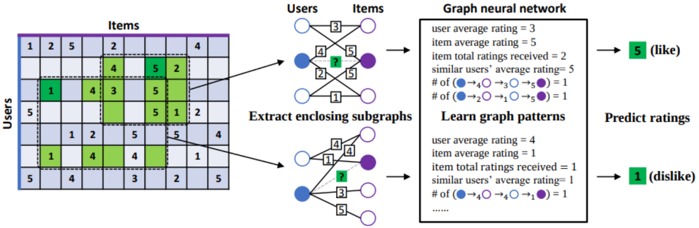
Pivo Recommendation



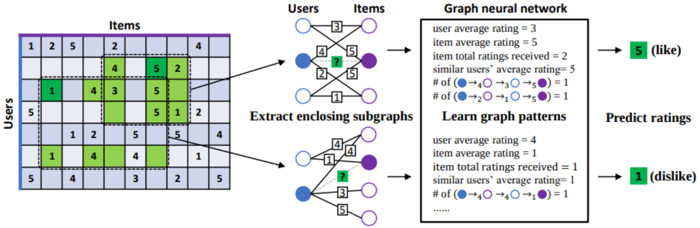
Addressing the Cold-Start Problem

- Recommender systems typically require millions of interactions
- However, new systems often have limited interaction data available
- Attribute-based recommendations can provide valuable information
  - Normally less significant than interactions themselves

Transfer Learning in RS



Transfer Learning in RS



Tandem (Marriage) Problem

1	C	B	E	A	D
2	A	B	E	C	D
3	D	C	B	A	E
4	A	C	D	B	E
5	A	B	D	E	C

A	3	5	2	1	4
B	5	2	1	4	3
C	4	3	5	1	2
D	1	2	3	4	5
E	2	3	4	1	5

Greedy Czech Algorithm

Rogue Combination

Rogue Combination

Gale & Shapley Algorithm

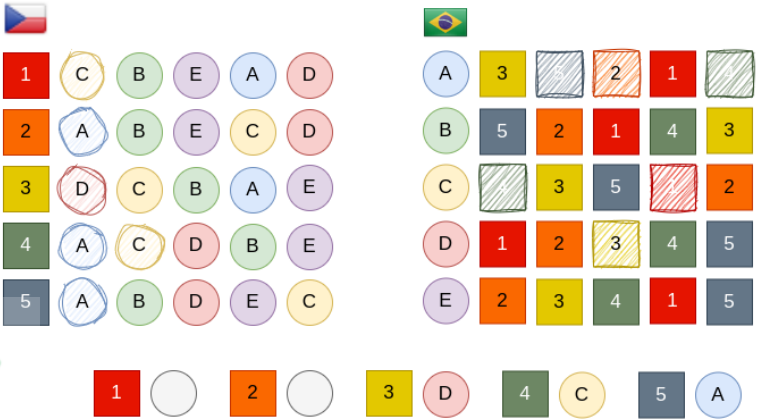
Gale & Shapley Algorithm

Gale & Shapley Algorithm

Gale & Shapley Algorithm

Gale & Shapley Algorithm

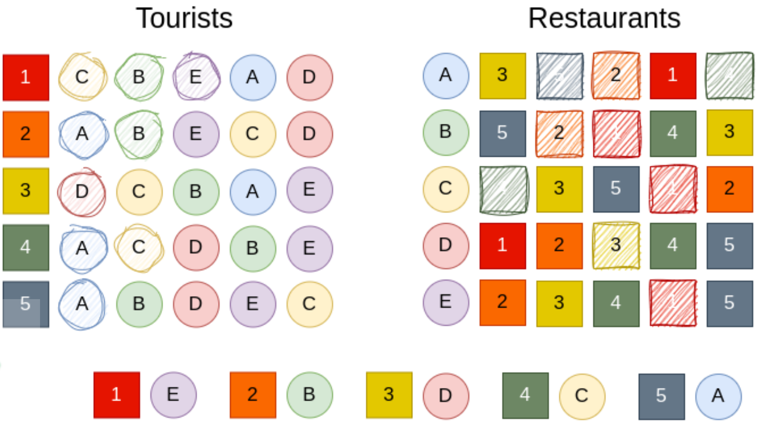
Gale & Shapley Algorithm



Gale & Shapley Algorithm...

What about Recommender Systems?

What about Recommender Systems?



Modern matters in RS

- Fairness
  - Recommender systems have the potential to perpetuate or even amplify bias
  - Unequal treatment of different groups of users
- Filter Bubbles
  - Common problem on RSs that rely heavily on personalization
  - Recommendations that align with a user's pre-existing preferences
  - Negative consequences for both individual users and society

Modern matters in RS

- Challenging to evaluate
  - Lack of ground truth
  - Changes over the time
  - Diversity of user preferences
- Scalability
  - When terabytes of memory is not enough
  - Can result in increased computational costs and reduced performance

Modern matters in RS

- Privacy concerns
  - Recommender systems often rely on user data to provide accurate recommendations
  - Legislation (GDPR)
  - Lack of interpretability
- Dynamic preference
  - User preferences and item characteristics can be highly dynamic
  - Item availability
  - Difficult to provide accurate and up-to-date recommendations

Obrigado :) - Faculty of Information Technology