# Modurec: Recommender Systems with Feature and Time Modulation

Javier Maroto, Clément Vignac, and Pascal Frossard

6-11 June 2021

EPFL

# Table of Contents

# Table of Contents

Goal: **Matrix reconstruction**

- Rows $\longrightarrow$ Users
- Columns $\longrightarrow$ Items
- Very few known values (less than 1% of entries)
- Many items with few users, few items with many users

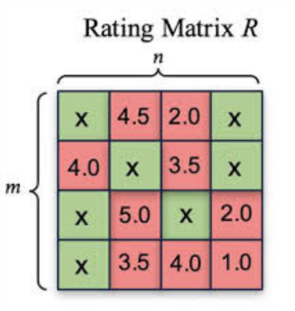



**Figure 1:** Matrix reconstruction problem

# Two main strategies

## Content filtering

- Recommend similar items to the ones the user liked
- Requires prior info on items (e.g. movie genre, lead actor...)
- This info provides limited information
- Performance does not scale with data

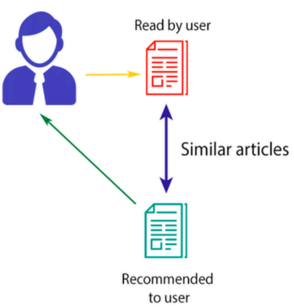



**Figure 2:** Content filtering

# Two main strategies

## Collaborative filtering

- Recommendation based on user with similar rating history
- Performance scales with data
- Rich information based on user behavior
- Susceptible to cold start and concept drift

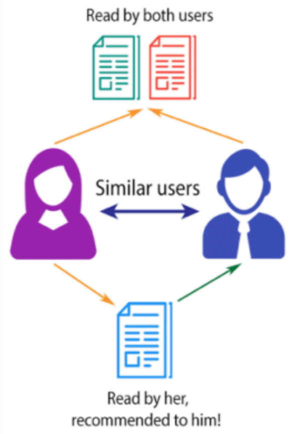



**Figure 3:** Collaborative filtering

# Cold start and concept drift

## Cold start

- Bad performance on new users or items due to lack of ratings
- Could stop new users from joining the platform

## Concept drift

- Bad performance on older users or items due to distribution shifts over time

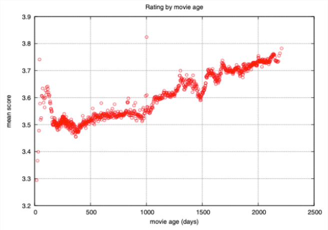



**Figure 4:** Ratings are time-dependent

# Table of Contents

# Our proposal

- Modurec: Address cold start and concept drift within a collaborative setting

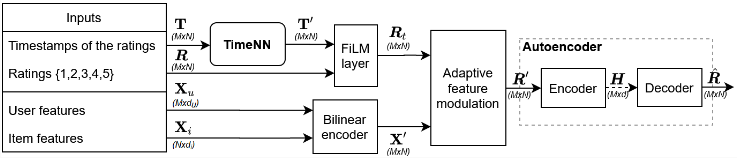
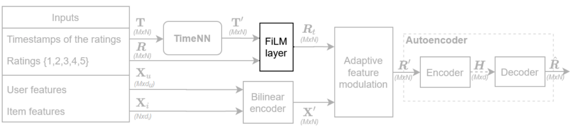



Figure 5: Modurec architecture

# Addressing concept drift: FiLM layers

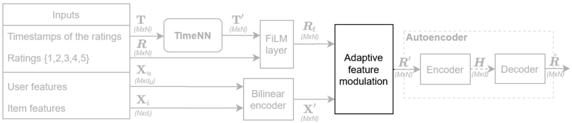



Feature-wise Linear Modulation (FiLM)

- Used to combine time and rating information
- Only 3 free parameters
- Much more expressive than concatenation

$$R_t = \alpha R + \beta T' + \gamma R \cdot T'$$

# Addressing cold start: adaptive feature modulation




## Adaptive feature modulation

- Adds the user/item feature information
- Uses a importance matrix to leverage when the feature information is most valuable (i.e. cold start)

$$A_{ij} = \begin{cases} \sigma(w_1|\mathcal{O}_{i,i}| + w_2|\mathcal{O}_{u,j}| + b) & |\mathcal{O}_{i,i}|, |\mathcal{O}_{u,j}| > 0 \\ 0 & |\mathcal{O}_{i,i}|, |\mathcal{O}_{u,j}| = 0 \end{cases}$$

$$R' = A \cdot R_t + (1 - A) \cdot X'$$

# Table of Contents

# Ablation study

**Table 1:** Average RMSE recommendation results on several datasets.
We use Modurec_[DFT] as the nomenclature for our model.
D = with dropout; F = with user and item features module; T = with time module.

| Dataset | GRALS | sRGCNN | GC-MC | STAR -GCN | CF-NADE | Sparse FC | TimeSVD++ flipped* | I-Autorec* | Modurec _D | Modurec _DT | Modurec _DFT |
|---|---|---|---|---|---|---|---|---|---|---|---|
| ML-100K | 0.945 | 0.929 | 0.905 | 0.895 | — | — | 0.890 | 0.908 | 0.905 | **0.887** | **0.887** |
| ML-1M | — | — | 0.832 | 0.832 | 0.829 | 0.824 | 0.842 | 0.831 | 0.826 | **0.821** | **0.821** |
| ML-10M | — | — | 0.777 | 0.770 | 0.771 | 0.769 | **0.749** | 0.782 | 0.789 | 0.777 | 0.779 |

- The **FiLM layer** that adds the time information has the biggest impact in performance
- The **adaptive feature modulation** does not seem to impact performance ⟶ very few test ratings with cold start

# Cold start evaluation

**Table 2:** Effect of combining the user and item features of the *Static* ($A_{ij} = \alpha$) and the *Adaptive* algorithms versus the *Nothing* algorithm, for users and items that have “few ratings” or “many ratings”.

| Dataset | Algorithm | Few ratings | Many ratings |
|---|---|---|---|
| ML-100K | Nothing | 1.6093 | **0.8371** |
| | Static | 1.6000 | 0.8417 |
| | Adaptive | **1.3412** | 0.8380 |
| ML-1M | Nothing | 1.1481 | **0.7895** |
| | Static | 1.1457 | 0.7900 |
| | Adaptive | **1.1360** | 0.7897 |

- The **adaptive feature modulation** improves a lot the performance on cold start scenarios
- Does not hinder performance on the rest of cases (unlike *Static*)

# Table of Contents

# Conclusion

- Data-driven approaches should be given more importance
- If the information is complemenatry (e.g. time), FiLM layers can be useful
- If the information is conflictive (e.g. feature vs rating information), adaptive schemes can be useful