

Group recommendation inspector

Real MovieLens (latest-small) · groups from `gr.groups.synthetic` (similar / divergent / outlier) with consensus-derived interactions · the slider is each member's **importance weight** (normalised to sum to 100%)

Inspector

How it's built

GroupIM [E2E]

1

Group 5 (divergent) · users 320, 367, 445

2

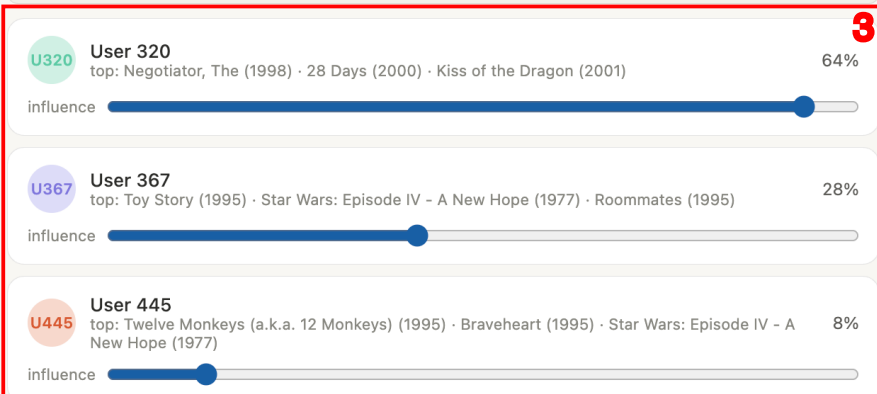
members: divergent ($r=0.03$)

Reset weights

Deep model encoding each member's history, pooled by an attention aggregator; a learned predictor scores items. Influence reweights the aggregator (real forward pass).

GROUP MEMBERS

GroupIM (deep): the influence weight is injected into the model's own attention aggregator and precomputed on a grid — NOT SAE steering. Each step is a real forward pass; equal weights = the native model.



ATTRIBUTION VIEW

Item history

Recs received

Latent concepts

User 320:

Comedy / Drama [Family Man, The (2000), All About Eve (1950), Holiday Inn (1942)]

Comedy / Drama [Larger Than Life (1996), Beavis and Butt-Head Do America (1996), Muppet Treasure Island (1996)]

Drama / Thriller [Crimson Tide (1995), Net, The (1995), Apollo 13 (1995)]

User 367:

Comedy / Drama [Family Man, The (2000), All About Eve (1950), Holiday Inn (1942)]

Comedy / Drama [Larger Than Life (1996), Beavis and Butt-Head Do America (1996), Muppet Treasure Island (1996)]

Drama / Thriller [Crimson Tide (1995), Net, The (1995), Apollo 13 (1995)]

User 445:

Comedy / Drama [Family Man, The (2000), All About Eve (1950), Holiday Inn (1942)]

Comedy / Drama [Larger Than Life (1996), Beavis and Butt-Head Do America (1996), Muppet Treasure Island (1996)]

Drama / Thriller [Crimson Tide (1995), Net, The (1995), Apollo 13 (1995)]

Concepts a member triggers (top-3 by activation over their history), from a Top-K SAE over GroupIM's item embeddings. These explain member taste profiles — the SAE is not involved in the model's attention-based steering.

● User 320 ● User 367 ● User 445

GROUP RECOMMENDATIONS (TOP 5 OF 50 CANDIDATES) · number = group relevance score (0–1)

1	Silence of the Lambs, The (1991) ✓ consensus item	1.00		
2	American President, The (1995) candidate	0.64		
3	Beneath the Planet of the Apes (1970) candidate	0.29		
4	Wild, The (2006) candidate	0.04		
5	Sully (2016) candidate	0.03		

Number = group relevance score (0–1) · dots = each user's share of that item's relevance (hover for %) · amber ✓ = the group's consensus item (rated ≥ 4 by ≥ 2 members)

AGGREGATE USER CONTRIBUTION ACROSS THE TOP 5

