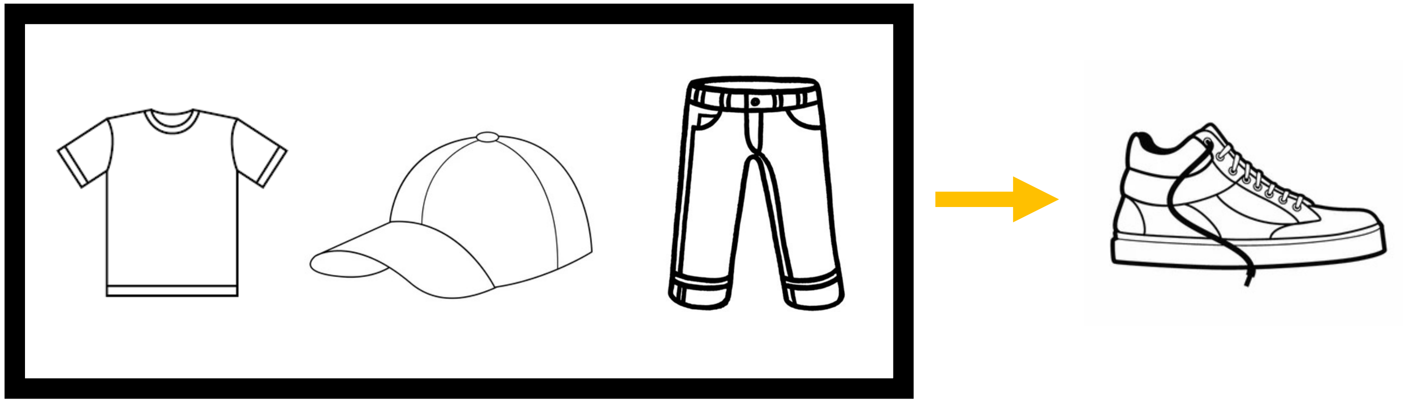


ADER: Adaptively Distilled Exemplar Replay Towards Continual Learning for Session-based Recommendation

Presentation by Alec Pauli

Session-based Recommendation Systems

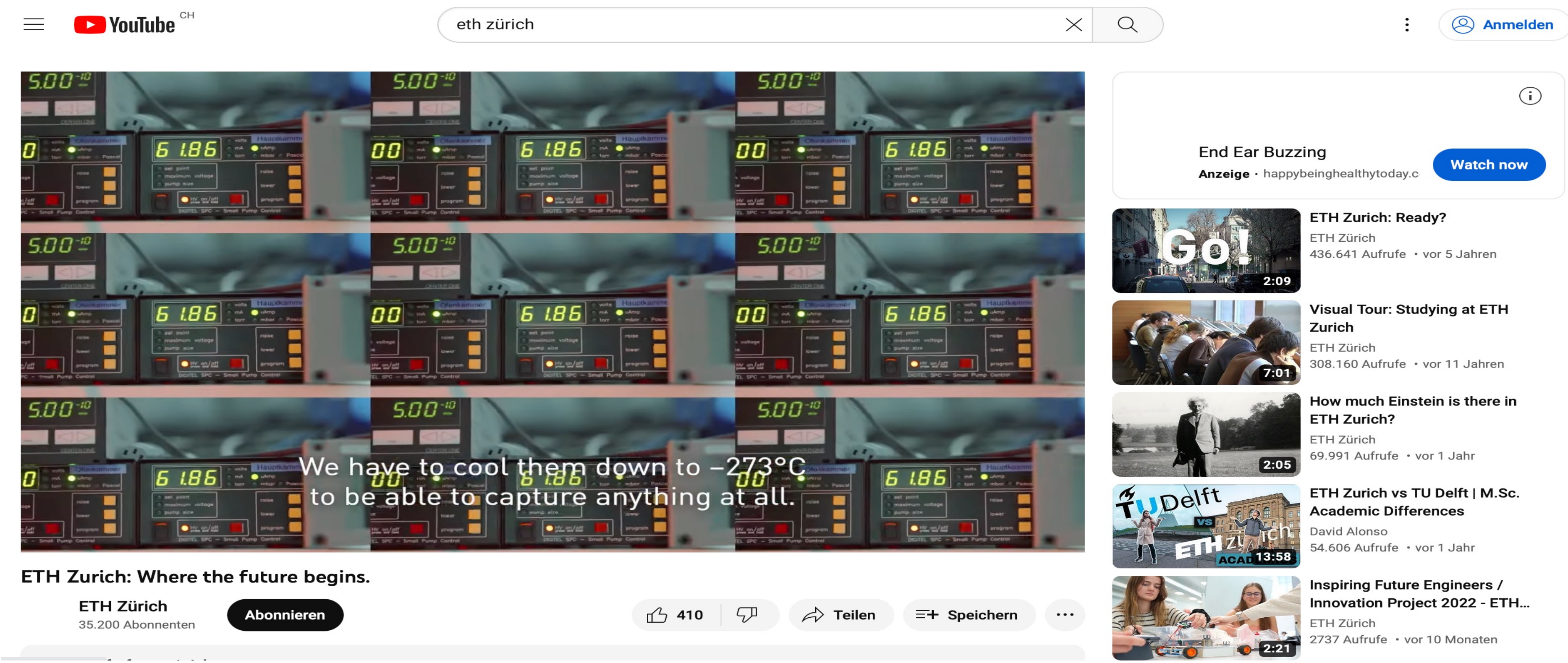
Recommender Systems

YouTube:
- 168,055,344,000 hours of video
- Enough to watch for more than 200'000 lives

Amazon:
- 350 million products
- Enough to buy every day around 12000 products

* Based on a life expectancy of 80 years

YouTube:


YouTube CH
eth zürich
Anmelden
We have to cool them down to –273°C
to be able to capture anything at all.
ETH Zurich: Where the future begins.
ETH Zürich
35.200 Abonnenten
Abonnieren
410
Teilen
Speichern
End Ear Buzzing
Anzeige · happybeinghealthytoday.c
Watch now
Go!
2:09
ETH Zurich: Ready?
ETH Zürich
436.641 Aufrufe · vor 5 Jahren
7:01
Visual Tour: Studying at ETH Zurich
ETH Zürich
308.160 Aufrufe · vor 11 Jahren
2:05
How much Einstein is there in ETH Zurich?
ETH Zürich
69.991 Aufrufe · vor 1 Jahr
13:58
ETH Zurich vs TU Delft | M.Sc. Academic Differences
David Alonso
54.606 Aufrufe · vor 1 Jahr
2:21
Inspiring Future Engineers / Innovation Project 2022 - ETH...
ETH Zürich
2737 Aufrufe · vor 10 Monaten

Amazon

Kunden, die diesen Artikel angesehen haben, haben auch angesehen

ASUS GeForce Dual RTX 3060 12GB V2 OC Edition Gaming Grafikkarte (GDDR6 Speicher, PCIe 4.0, 1x…

1.394

337,42 €

Erhalte es bis **Dienstag, 9. Mai**

GRATIS-Versand für Bestellungen ab 49,00 € und Versand durch Amazon

ASUS Dual Nvidia GeForce RTX 2060 6GB EVO OC Edition Gaming Grafikkarte (GDDR6 Speicher, PCIe 3.0, 1x…

1.527

249,15 €

Erhalte es bis **Dienstag, 9. Mai**

GRATIS-Versand für Bestellungen ab 49,00 € und Versand durch Amazon

ASUS Phoenix GeForce RTX 3050 8G Gaming Grafikkarte (NVIDIA Ampere, 8GB GDDR6 Speicher, PCIe 4.0, 1x…

123

280,64 €

Erhalte es bis **Dienstag, 9. Mai**

GRATIS-Versand für Bestellungen ab 49,00 € und Versand durch Amazon

Nur noch 13 auf Lager

Zotac RTX 3050 TwinEdgeOC ZT-A30500H-10M

216

Spare 6%

259,98 €

Statt: ~~276,47 €~~

Erhalte es bis **Donnerstag, 11. Mai**

GRATIS-Versand für Bestellungen ab 49,00 € und Versand durch Amazon

ASUS TUF NVIDIA GeForce GTX 1660 TI 6G OC Edition Gaming Grafikkarte (PCIe 3.0, 6GB GDDR6 Speicher,…

200

243,25 €

Erhalte es bis **Mittwoch, 10. Mai**

GRATIS-Versand für Bestellungen ab 49,00 € und Versand durch Amazon

MSI GeForce RTX 3050 Ventus 2X 8G Gaming Grafikkarte - NVIDIA RTX 3050, 8 GB DDR6 Speicher, V397-435R,…

353

266,72 €

Erhalte es bis **Donnerstag, 11. Mai**

GRATIS-Versand für Bestellungen ab 49,00 € und Versand durch Amazon

Gigabyte GeForce RTX 3050 Eagle OC 8G NVIDIA 8 GB GDDR6

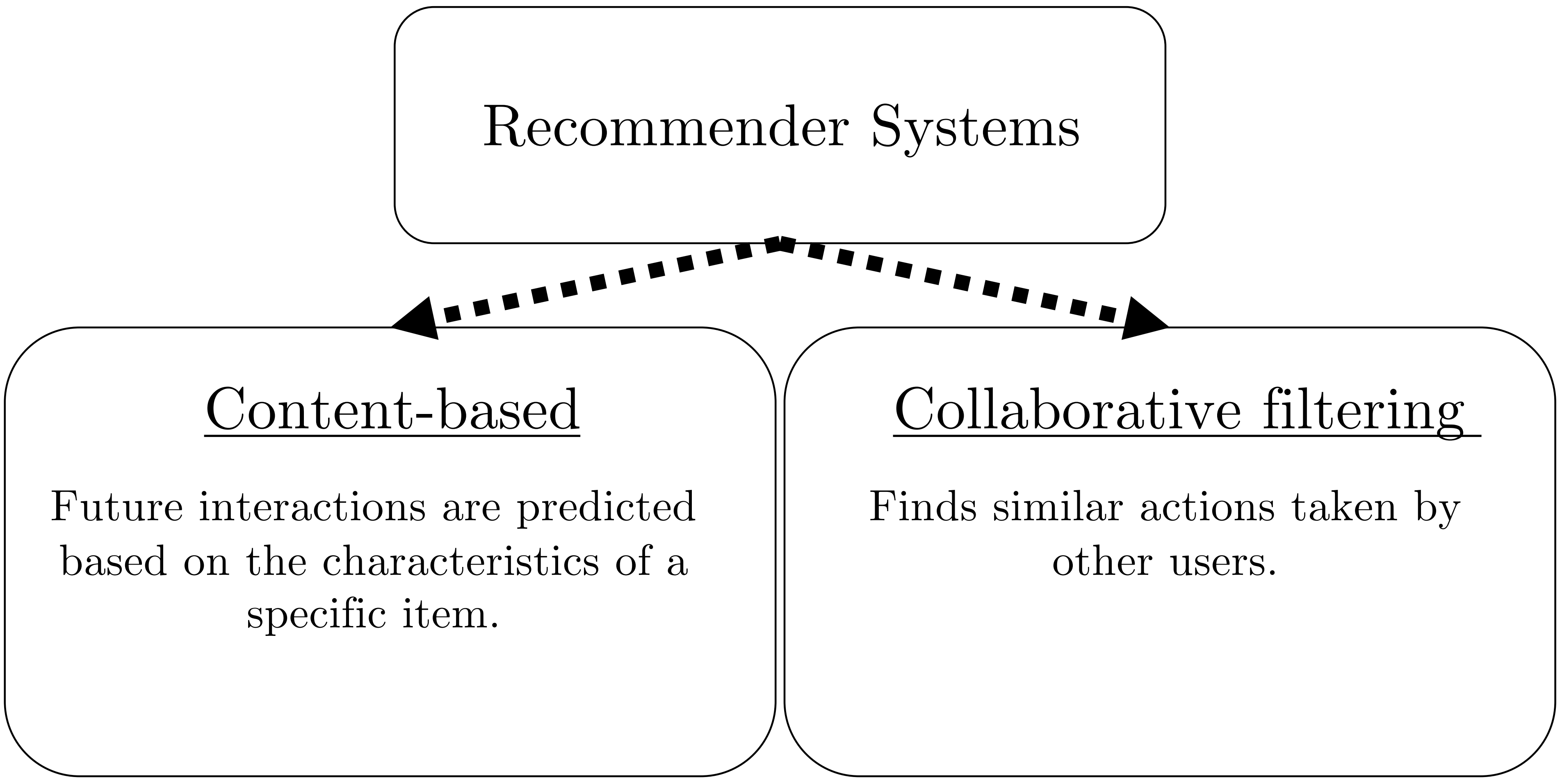
72

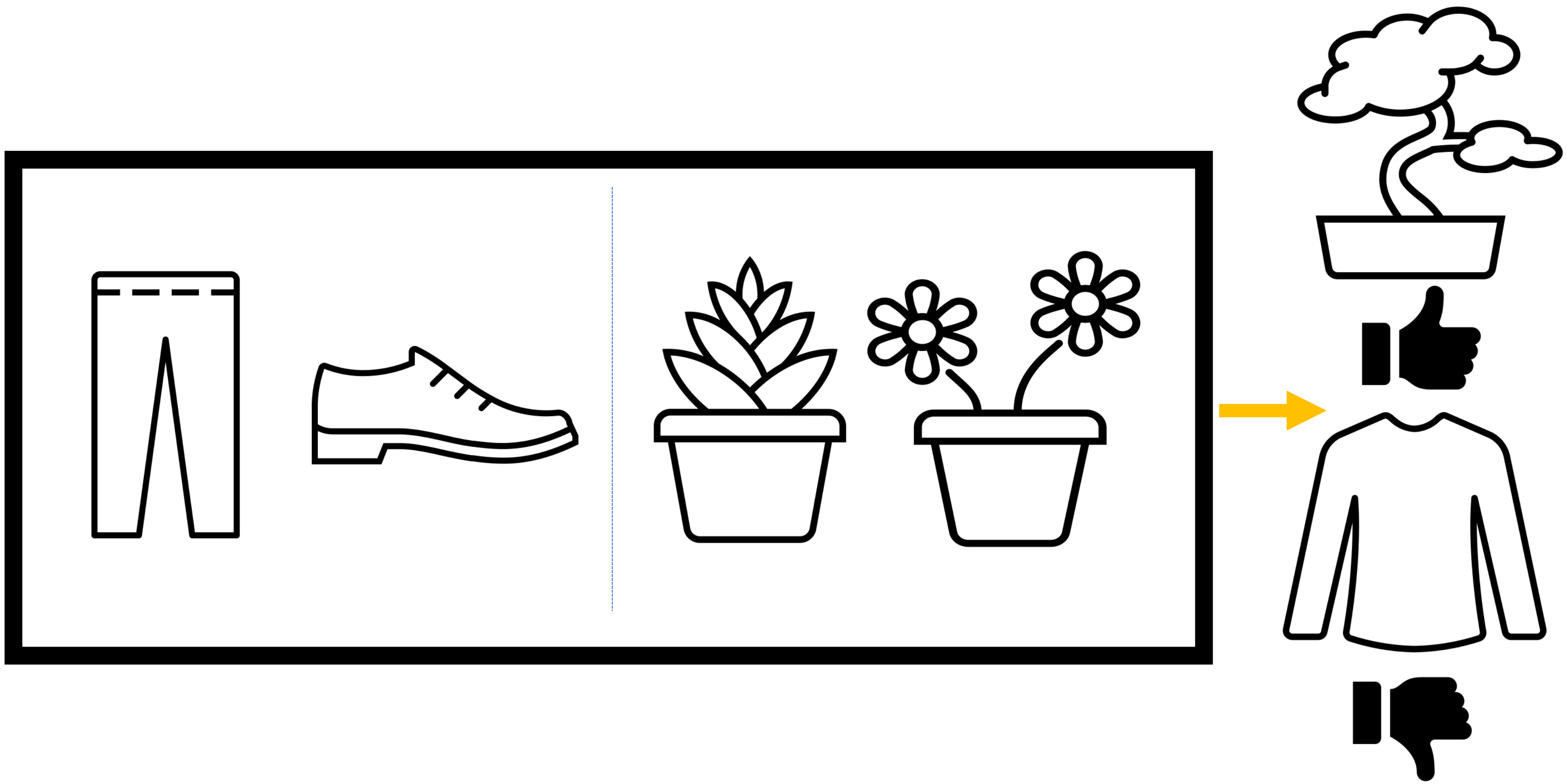
283,24 €

GRATIS-Versand für Bestellungen ab 49,00 € und Versand durch Amazon

Amazon

514 Billions * 1% * 10% = 0.5 Billion


Recommender Systems
Content-based
Future interactions are predicted based on the characteristics of a specific item.
Collaborative filtering
Finds similar actions taken by other users.



Session based recommender Systems

No login required

As a session is normally a short continuous interaction with the service already are able to predict without a large history

Privacy regulations

For example, GDPR makes storing large datasets more complicated. Even more important is the new E-Privacy regulation

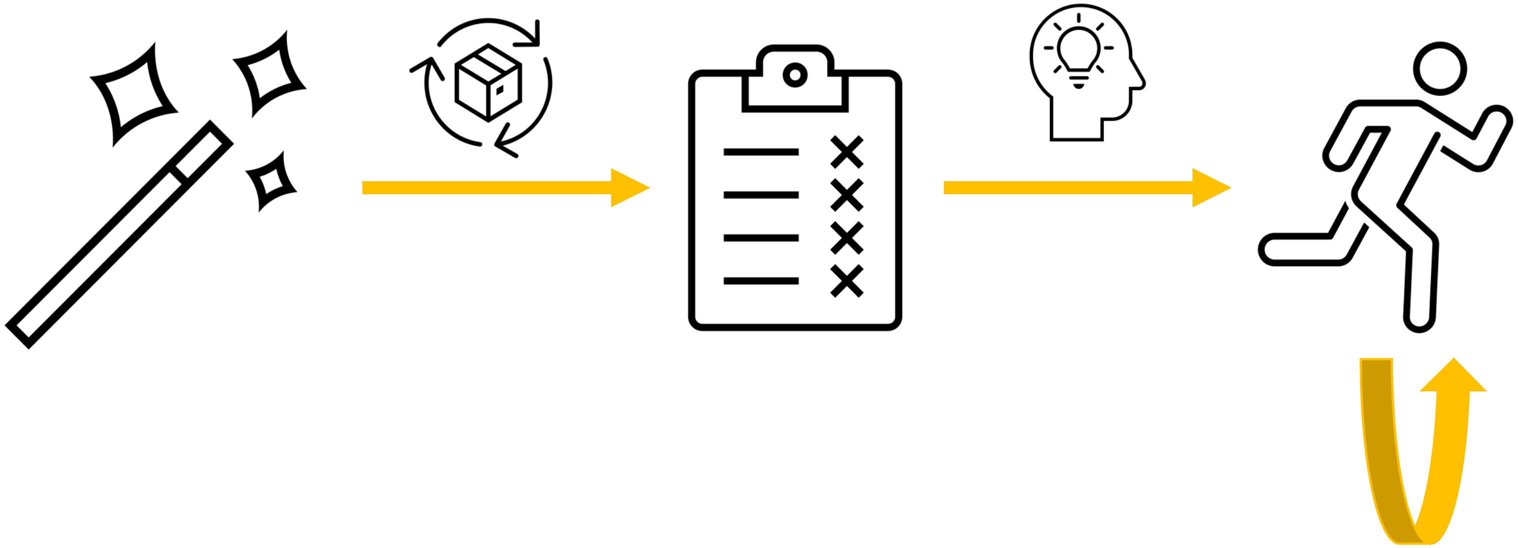
High quality datasets

With the rising popularity of new social networks and content platforms new datasets were open-sourced/collected

Deep Learning advances

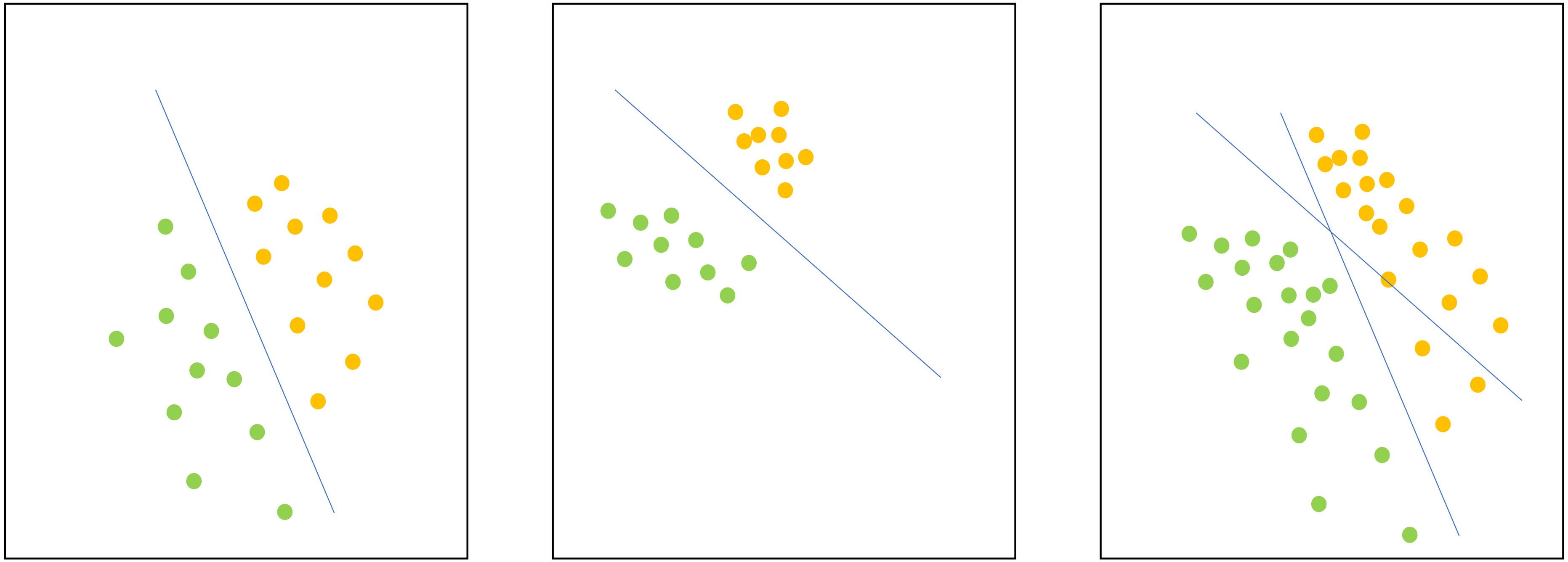
Naturally fits into paradigms of Deep learning. Thus advances in Deep Learning such as RNNs can be directly applied

Recommender Systems are dynamic system



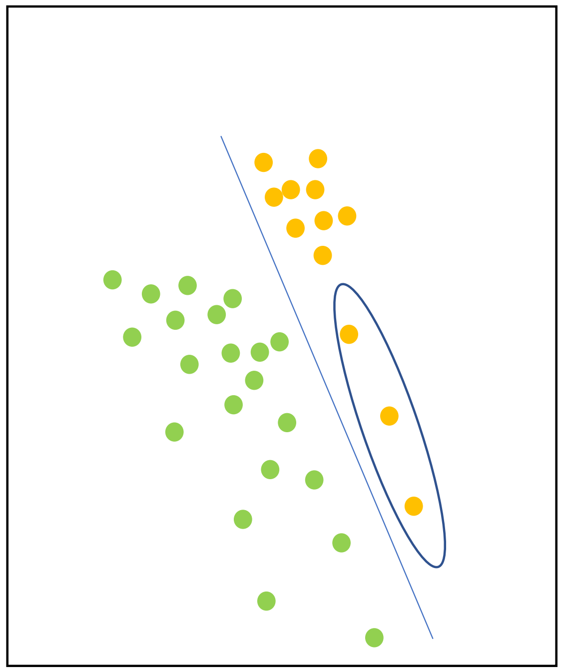
Catastrophic forgetting

Learning task 1 | Learning task 2 | Learning task f



Catastrophic cartoon solution

Learning task 2

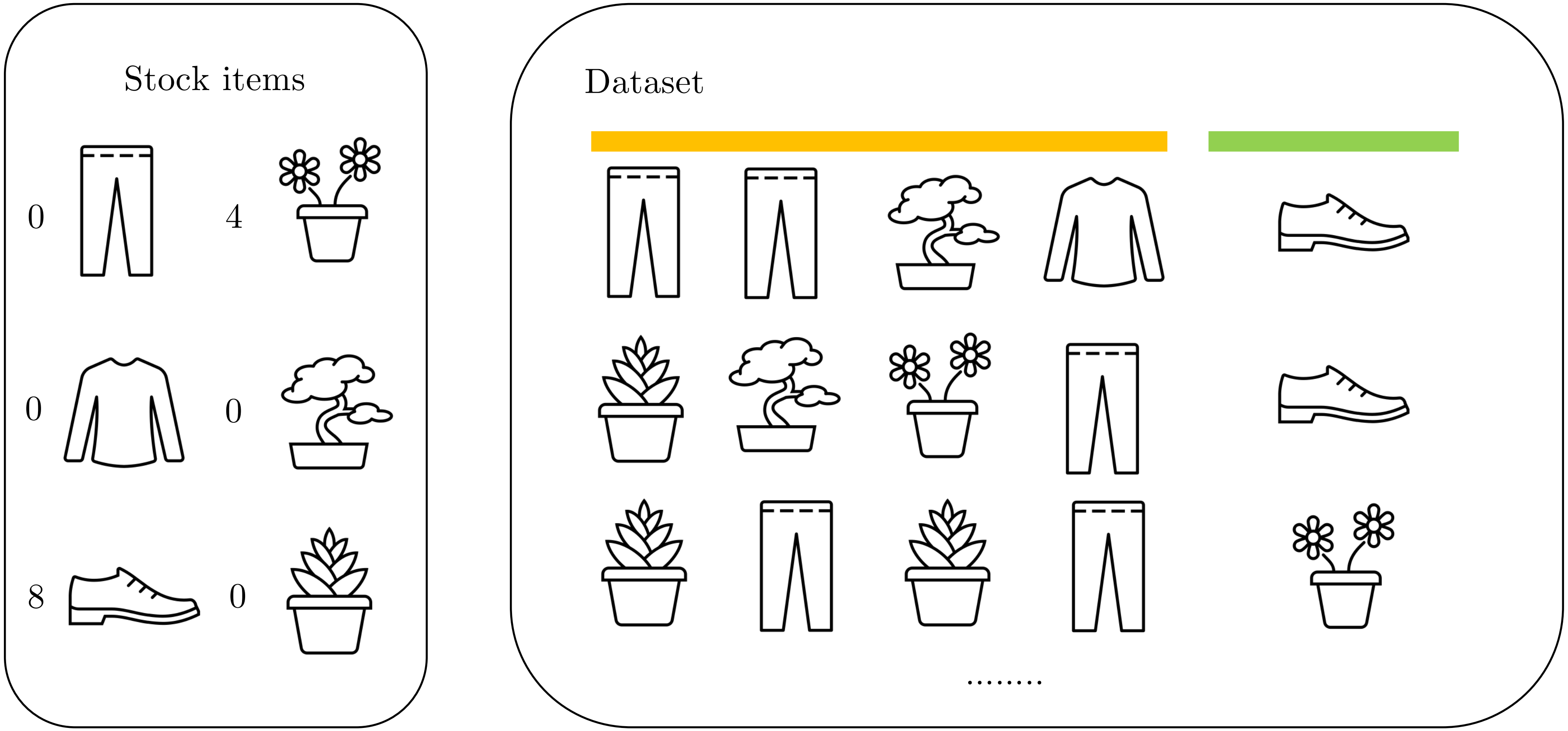


Ader

How to split available space ?

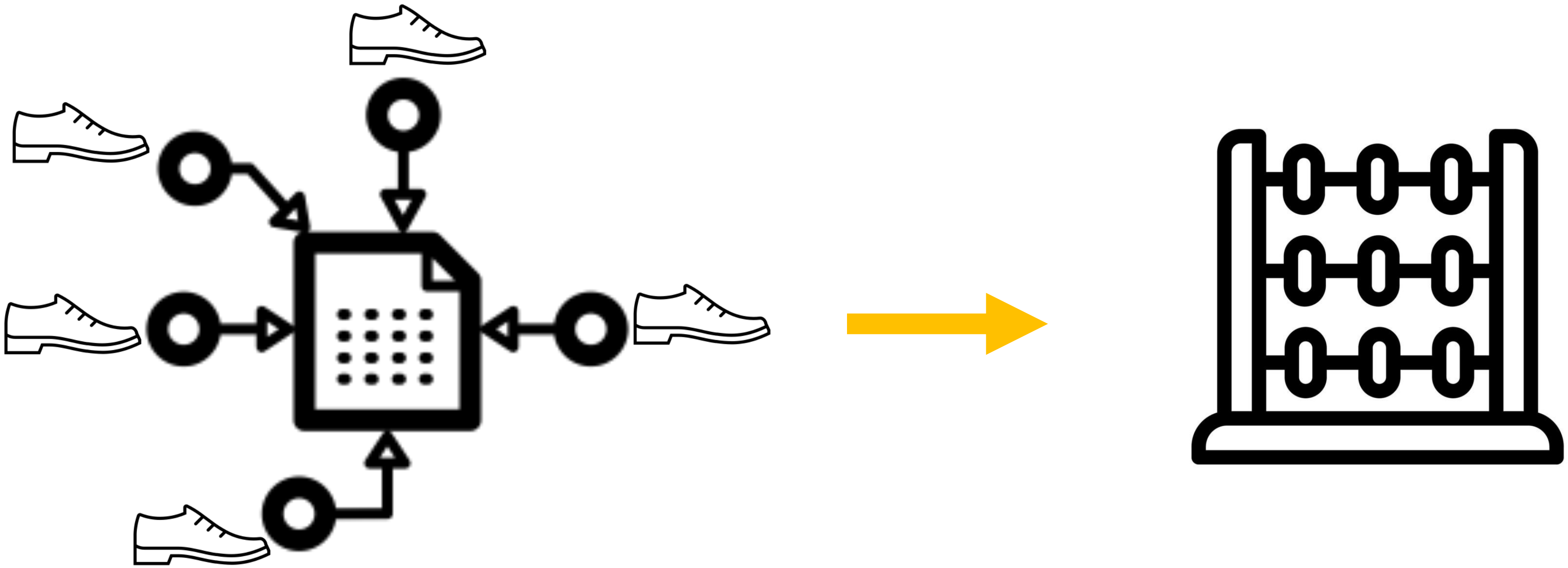
Which datapoints ?

Sample Dataset

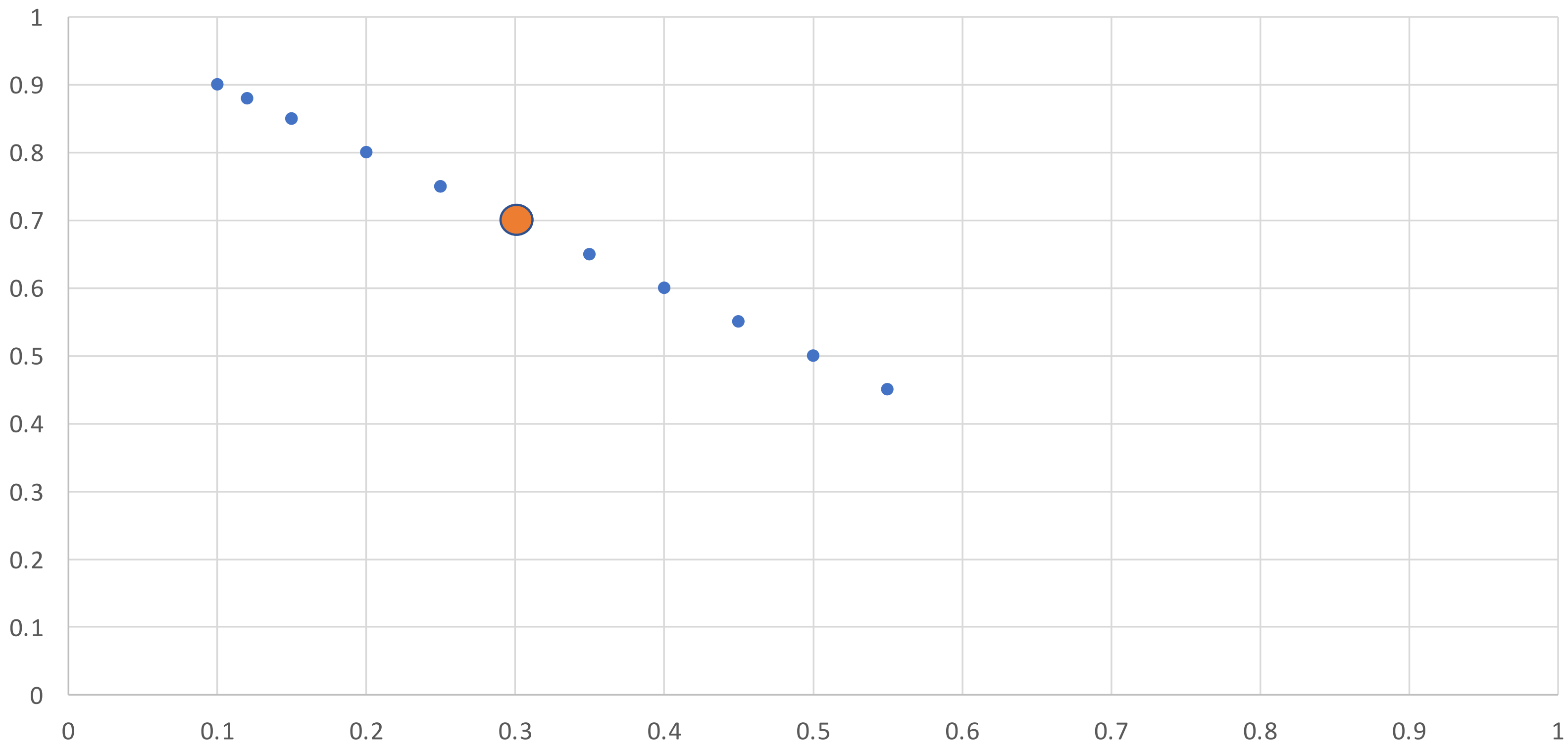


Split on the available space

Selection of Exemplars of the class



Selection of Exemplars of the  class



Open questions

How do we compute the loss for training ?

Which type of network is used ?

KD - Loss

CE - Loss

Combination

Model – SASRec – Previous models

Markov Chain

- Good at short term realtionships

Recurrent Neuronal Network

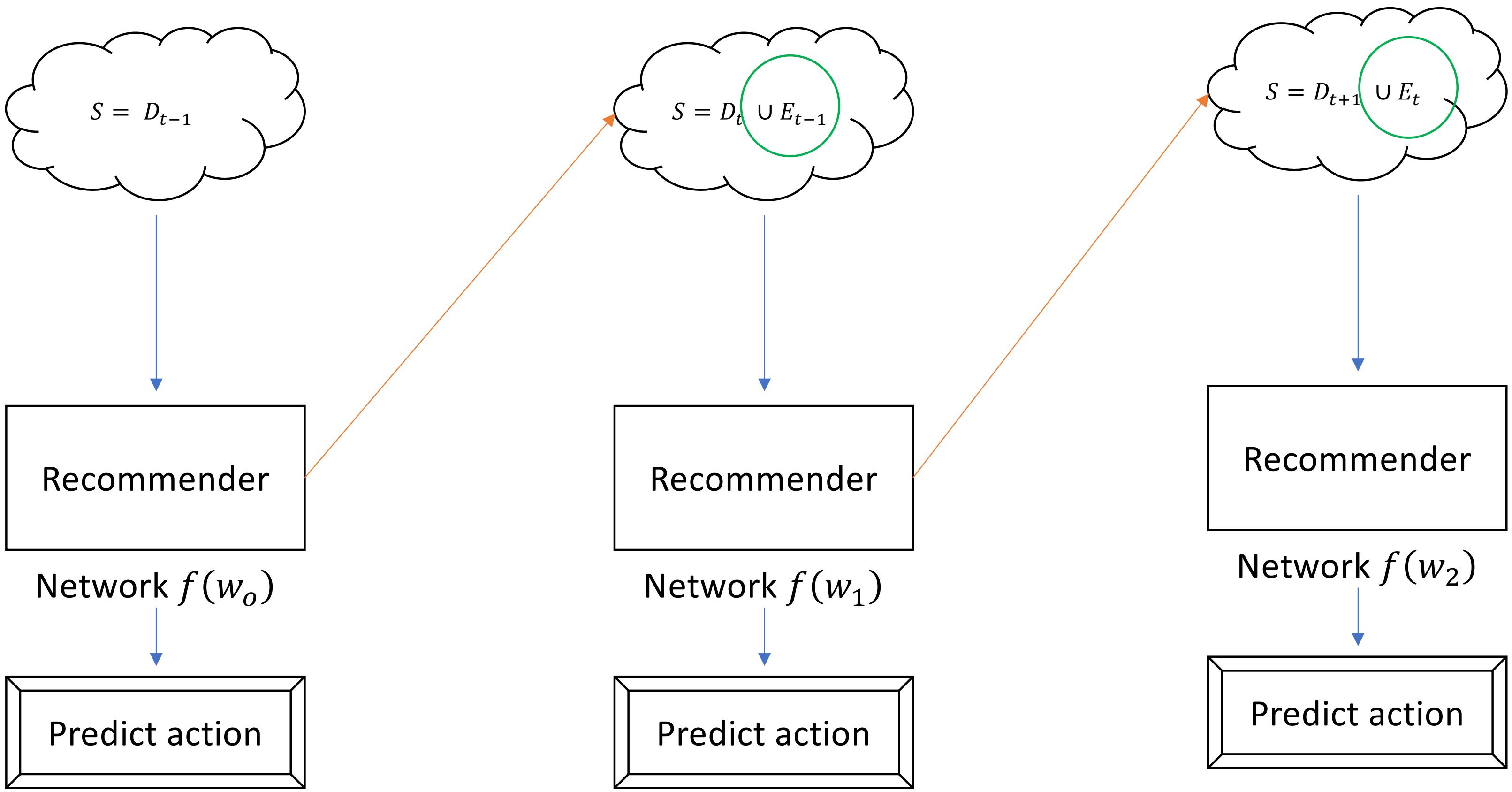
Perform best with long term semantics

Model – SASRec – High level idea

<u>SASRec</u>

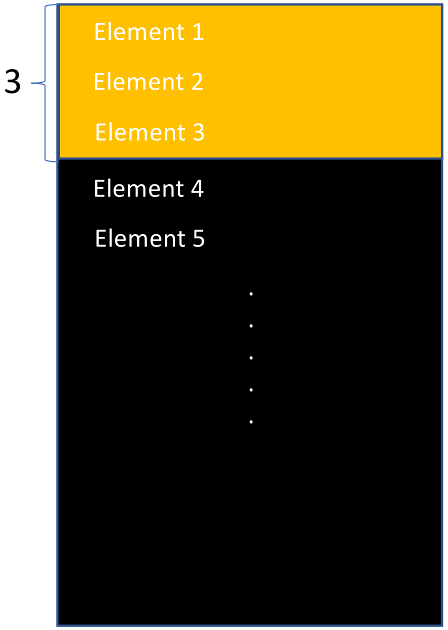
Tries to combine the strengths of MC and RNN's via an attention mechanism

Adaptively Distilled Exemplar Replay



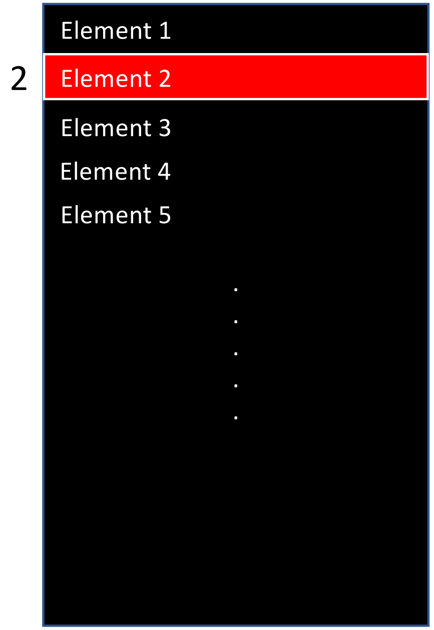
Experiments

Recall@k :



$$\frac{\#\text{in first k}}{\#\text{total}}$$

MRR@k :



$$\frac{1}{\#\text{total}} \sum_{i=1}^{\#\text{total}} \frac{1}{\text{corresp. m}}$$

Comparison against Ader

Finetune

Joint

Dropout

EWC

Dropout

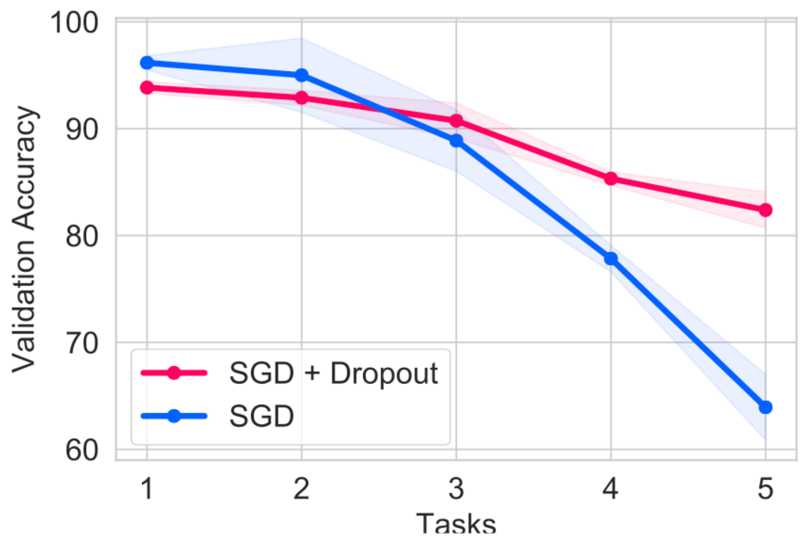


Figure 1. Networks trained with dropout tend to forget at a slower rate. The lines represent the evolution of the validation accuracy of the first task, as networks learn new tasks

EWC model

Diginetica

Click stream data of e-commerce site	5 Months

YouChoose

Click stream data of different e-commerce site

6 Months

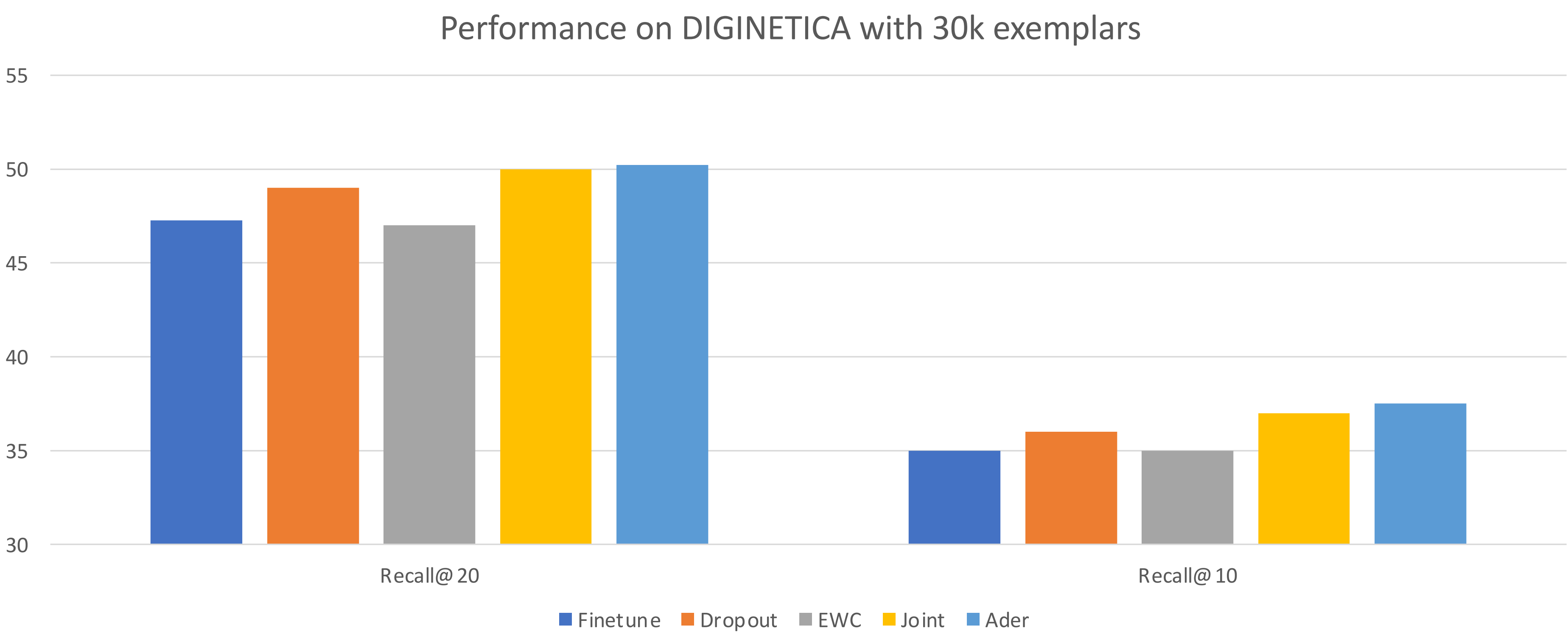
Less dynamic

YouChoose Diginetica

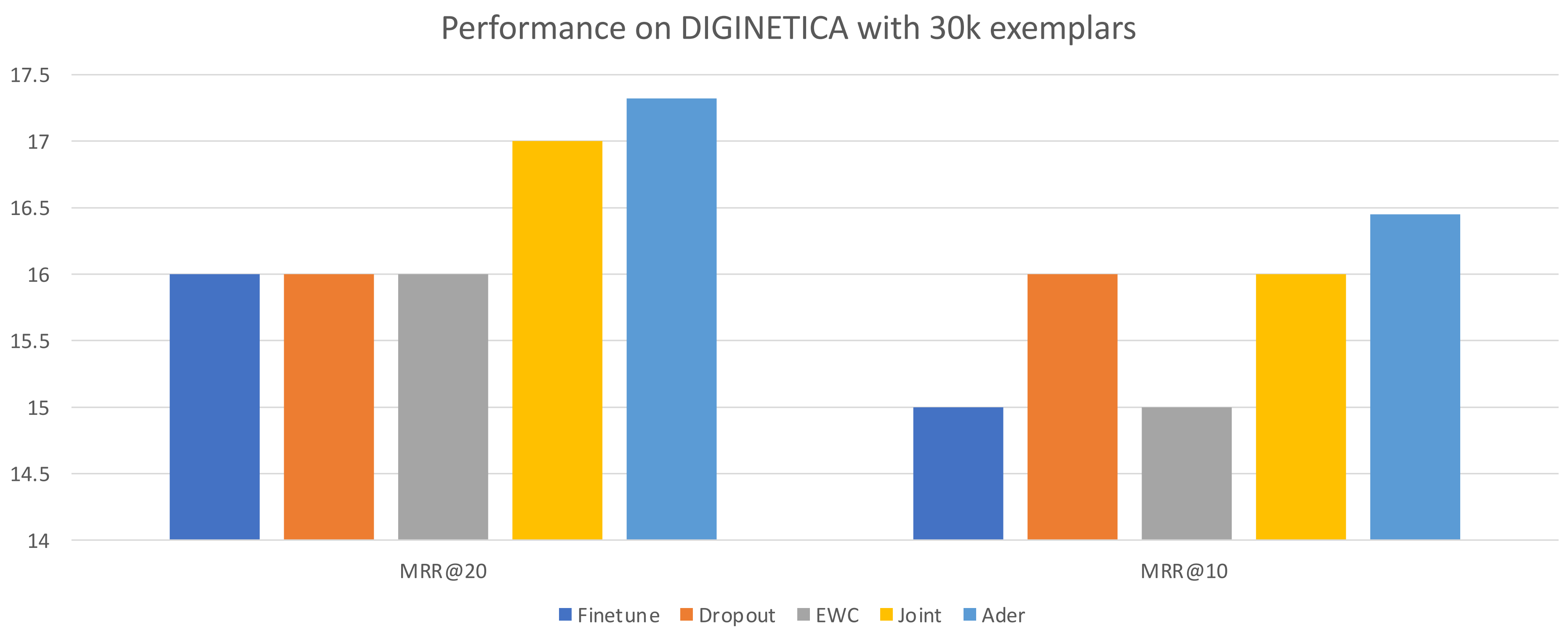
For YouChoose the update interval is daily and for Digitenica weekly

Still Youchoose has around 4 times more Actions in each intervall

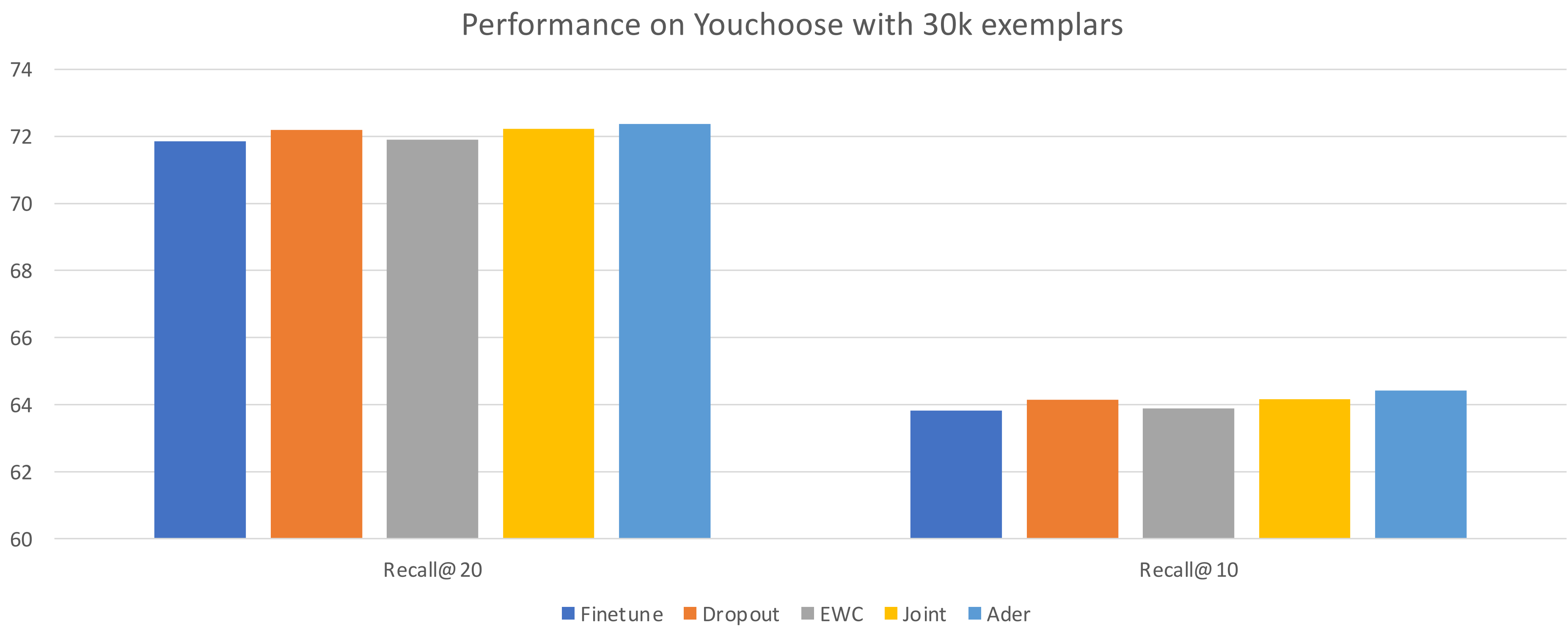
Results of the Diginetica dataset



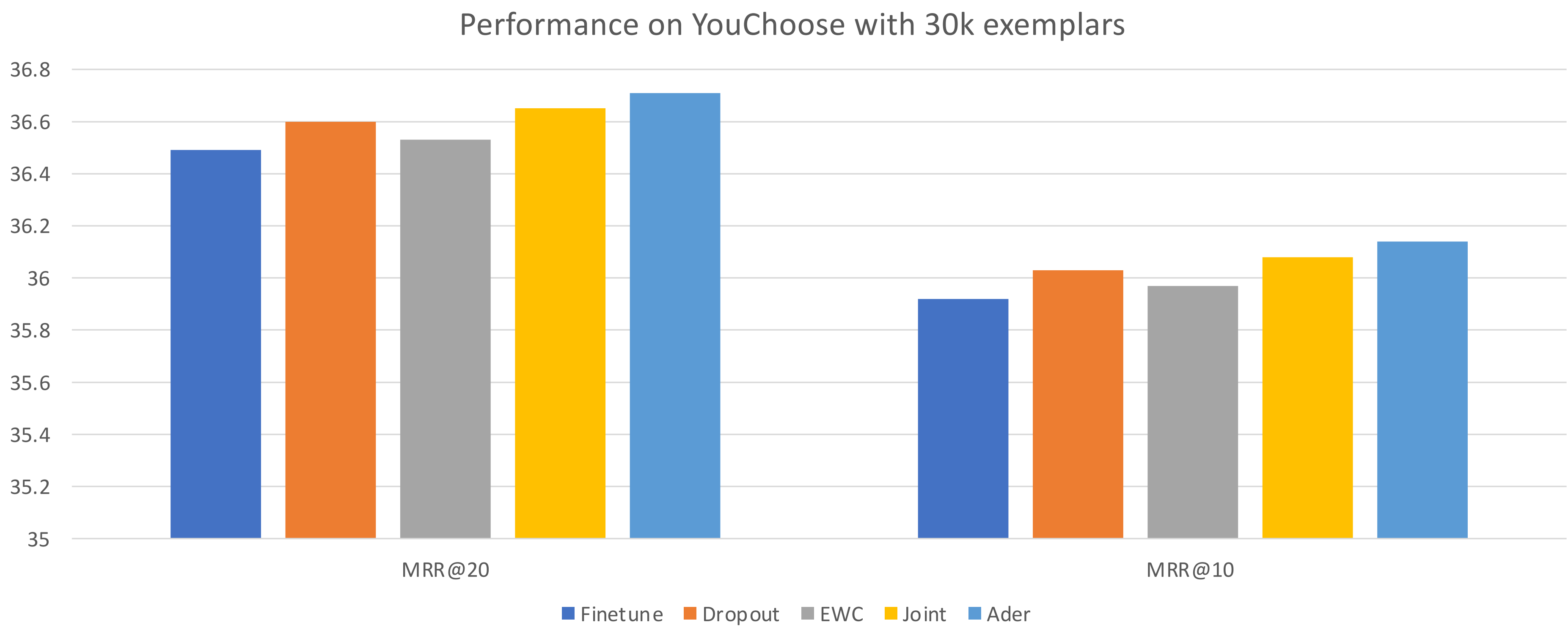
Results of the Diginetica dataset



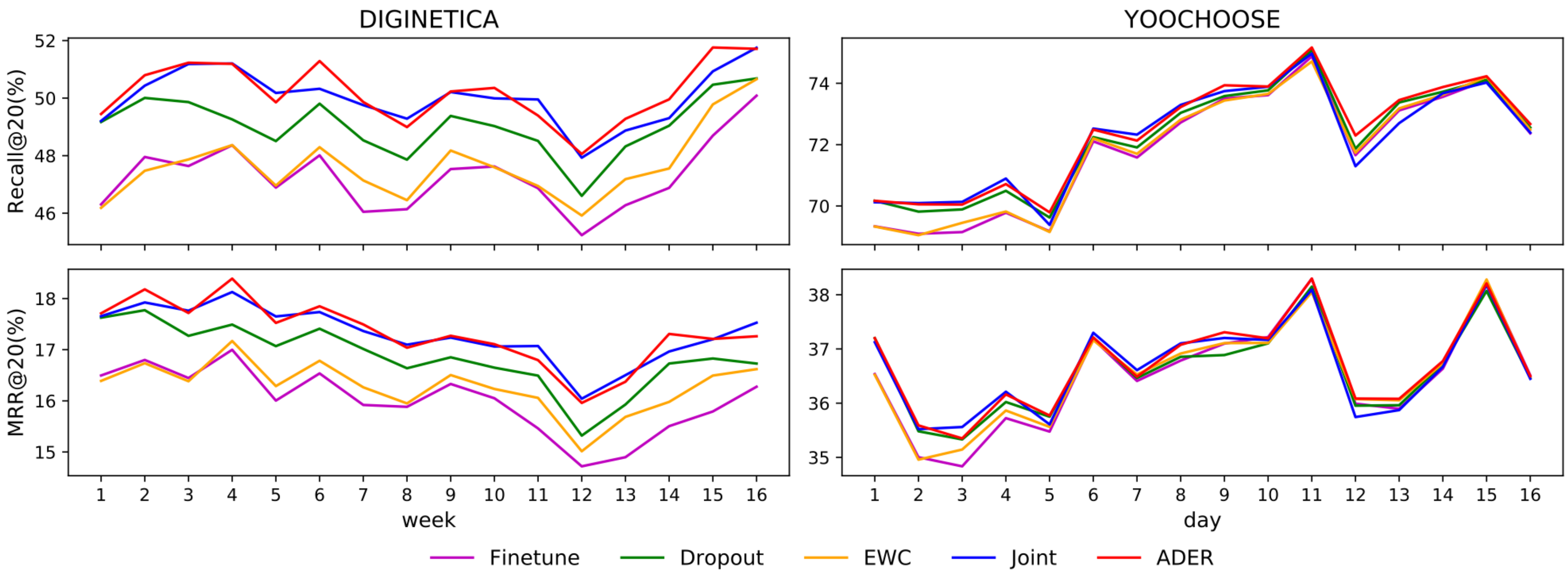
Results of the YOOCHOOSE dataset



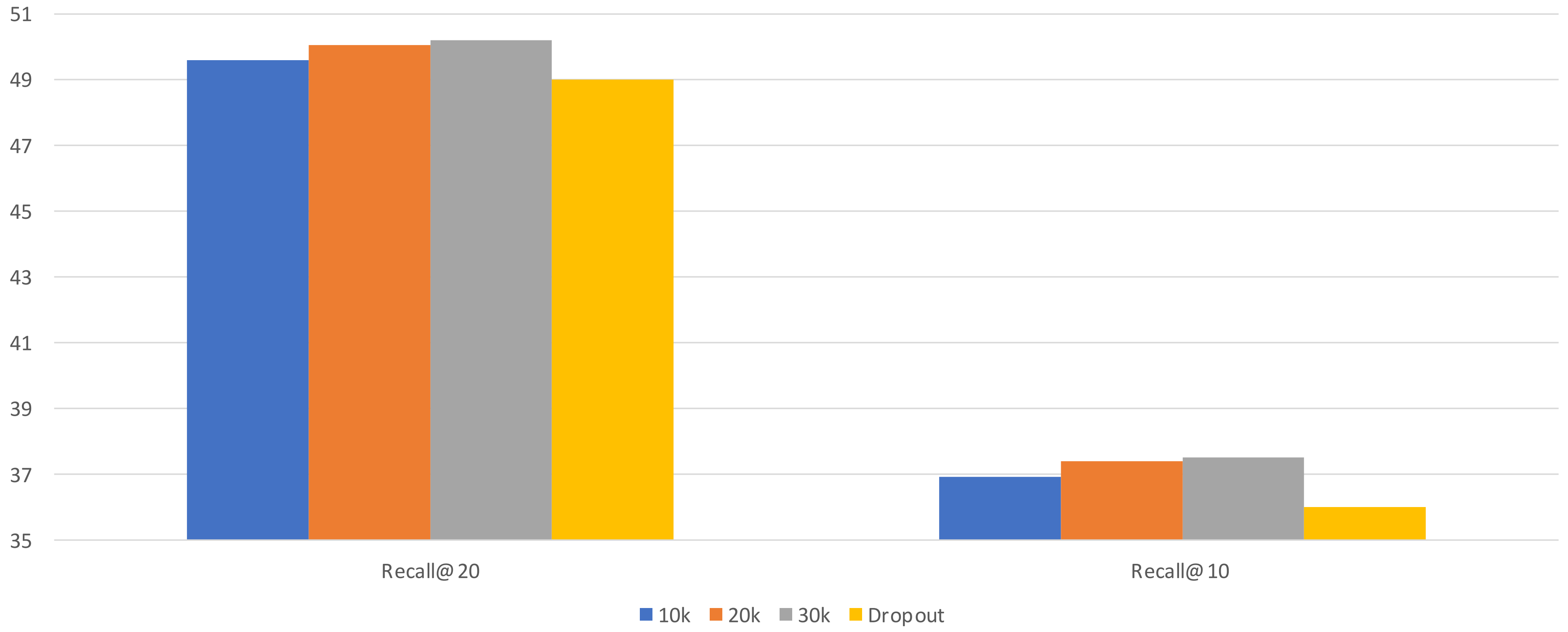
Results of the YOOCHOOSE dataset



Performance over the weeks



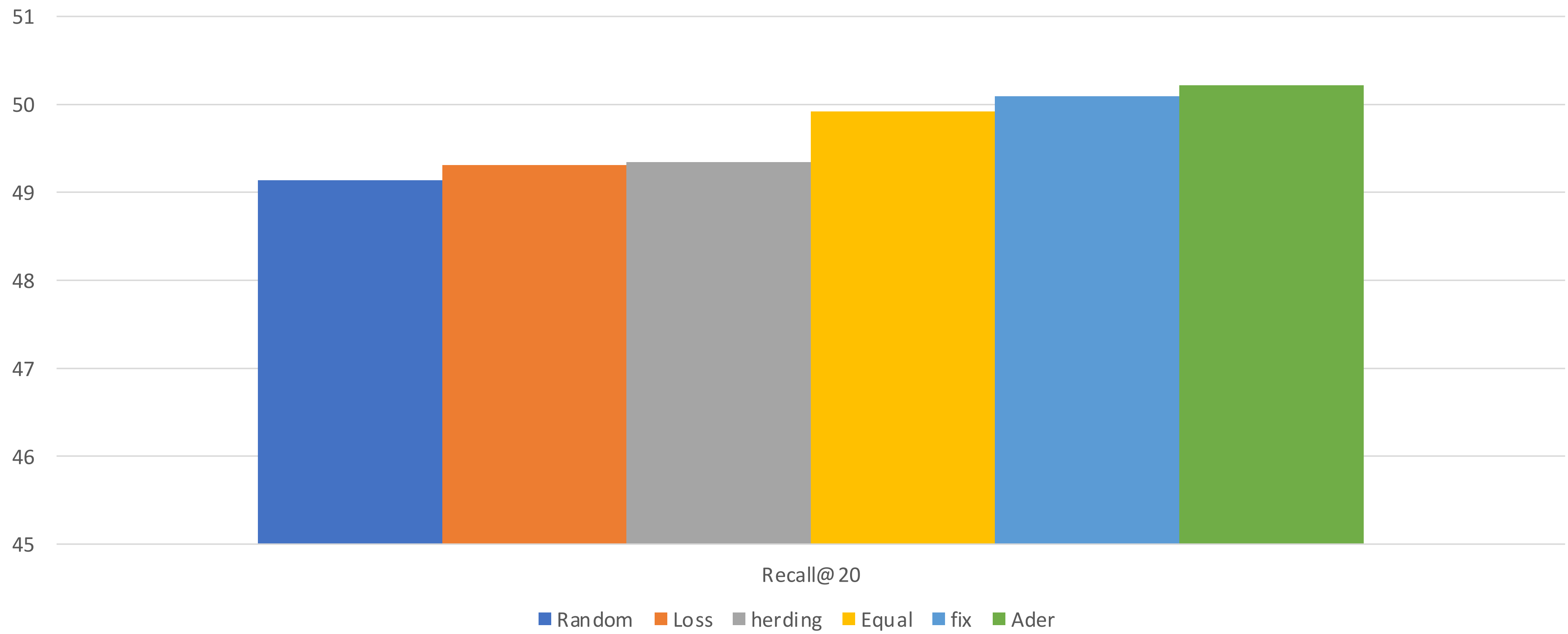
Effect of exemplar size



Ablation study

random	loss	herding
equal	fix	Ader

Ablation study



Personal opinion

<u>How to use space</u>

<u>Better than a upper baseline</u>

<u>Generally good written</u>

<u>Herding technique</u>

Exemplar sizes

Diginetica

Around 50'000 samples per iteration

YouChoose

Around 200'000 samples per iteration

Main sources:

- https://session-based-recommenders.fastforwardlabs.com/FF19-Session_Based_Recommender_Systems-Cloudera_Fast_Forward.pdf
- https://towardsdatascience.com/introduction-to-recommender-systems-1-971bd274f421
- https://medium.com/@mdsangha/session-based-recommendations-f16369aafa6bhttps://session-based-recommenders.fastforwardlabs.com/FF19-Session_Based_Recommender_Systems-Cloudera_Fast_Forward.pdf
- https://www.google.com/search?client=safari&rls=en&q=session+based+recommender+springer&ie=UTF-8&oe=UTF-8
- https://github.com/kang205/SASRec
- https://www.researchgate.net/publication/343179237_ADER_Adaptively_Distilled_Exemplar_Replay_Towards_Continual_Learning_for_Session-based_Recommendation
- https://www.google.com/url?sa=t&rct=j&q=&esrc=s&source=web&cd=&ved=2ahUKEwjwvpSn0-X-AhVDhP0HHSmCBQoQFnoECBMQAQ&url=https%3A%2F%2Farxiv.org%2Fpdf%2F1808.09781&usg=AOvVaw3a7U8EYJmlezuVqztD5Irj
- https://openaccess.thecvf.com/content_CVPRW_2020/papers/w15/Mirzadeh_Dropout_as_an_Implicit_Gating_Mechanism_For_Continual_Learning_CVPRW_2020_paper.pdf

Backup Slides / Support for discussion

CE Loss

Cross entropy according to current data

$$L_{CE}(\theta_t) = -\frac{1}{|D_t|} \sum_{(\mathbf{x}, y) \in D_t} \sum_{i=1}^{|I_t|} \delta_{i=y} \cdot log(p_i)$$

KD Loss

Softmax on all items

$$L_{KD}(\theta_t) = -\frac{1}{|E_{t-1}|} \sum_{(\mathbf{x},y)\in E_{t-1}} \sum_{i=1}^{|I_{t-1}|} \hat{p}_i \cdot log(p_i),$$

Old network

New network

Total Loss

Small if either a lot of new actions are available or a lot of new data

$$L_{ADER} = L_{CE} + \lambda_t \cdot L_{KD}, \quad \lambda_t = \lambda_{base} \sqrt{\frac{|I_{t-1}|}{|I_t|} \cdot \frac{|E_{t-1}|}{|D_t|}}$$

Algorithm for choosing exemplars

Pseudoalgorithm for selection in loop t:

For all items y:

$P_y =$ elements with the same y

$\mu =$ Average of the y according to the output of the model

for k from 1 to number of elements to store for this action

$$\mathrm{argmin}_{x \in P_y} \left| \mu - \frac{1}{k}\left(\phi(x) + \sum_{j=1}^{k-1} \phi(x_j)\right) \right|$$

Use the union of all elements chosen

Important training parameters

- SASRec used 150 hidden units and 2 stacked self-attention blocks
- Batch size is 256 for Diginetica and 512 for YOOCHOOSE
- The Adam Optimizier was used with a learning rate of 5e-4
- Train default was 100 epochs that were lowered if Recall@20 didn't improve for 5 epochs