Parallel Recurrent Neural Network Architectures for Feature-rich Session-based Recommendations

Balázs Hidasi, Gravity R&D (@balazshidasi)
Massimo Quadrana, Politecnico di Milano (@mxqdr)
Alexandros Karatzoglou, Telefonica Research (@alexk_z)
Domonkos Tikk, Gravity R&D (@domonkostikk)

GRAVITY
Research & Development

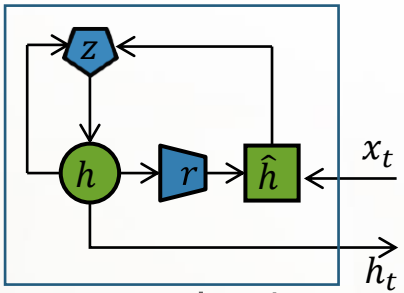
Session-based recommendations

- Permanent (user) cold-start
  - User identification
  - Changing goal/intent
  - Disjoint sessions
- Item-to-Session recommendations
  - Recommend to previous events
- Next event prediction

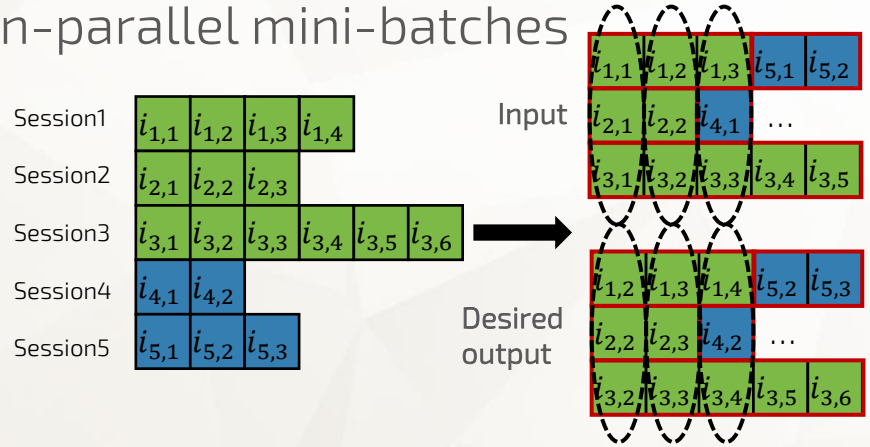
GRAVITY
Research & Development

GRU4Rec [Hidasi et. al. ICLR 2016]

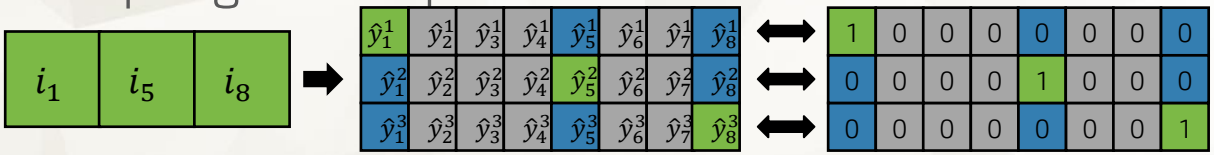
- Gated Recurrent Unit
  - $z_t = \sigma(W^z x_t + U^z h_{t-1})$
  - $r_t = \sigma(W^r x_t + U^r h_{t-1})$
  - $\hat{h}_t = \sigma(W x_t + U(r_t \circ h_{t-1}))$
  - $h_t = (1 - z_t) h_{t-1} + z_t \hat{h}_t$

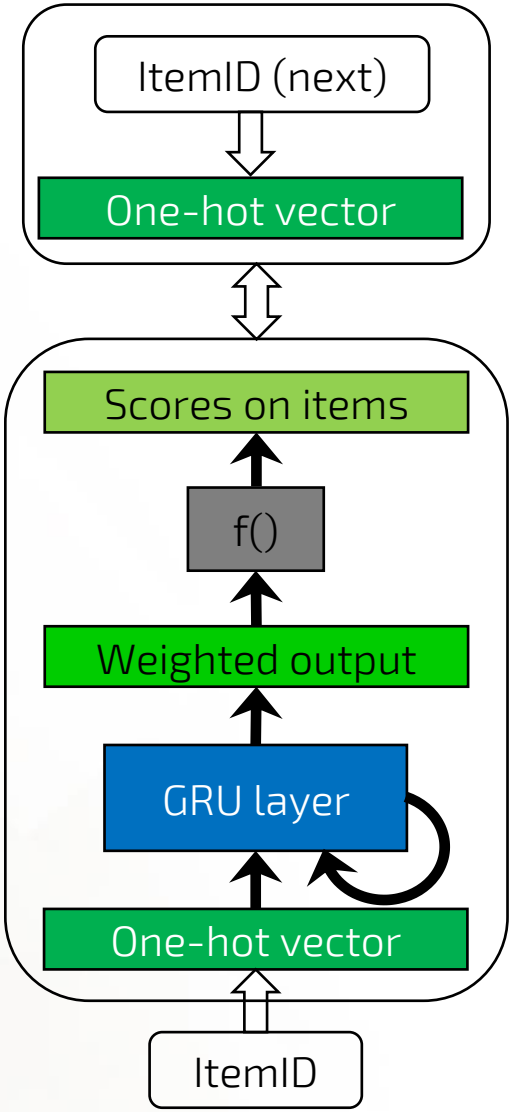


- Adapted to the recommendation problem
  - Session-parallel mini-batches



  - Ranking loss: TOP1= $\frac{1}{N_S}\sum_{j=1}^{N_S}\sigma(\hat{r}_{s,i} - \hat{r}_{s,j}) + \sigma(\hat{r}_{s,j}^2)$
  - Sampling the output





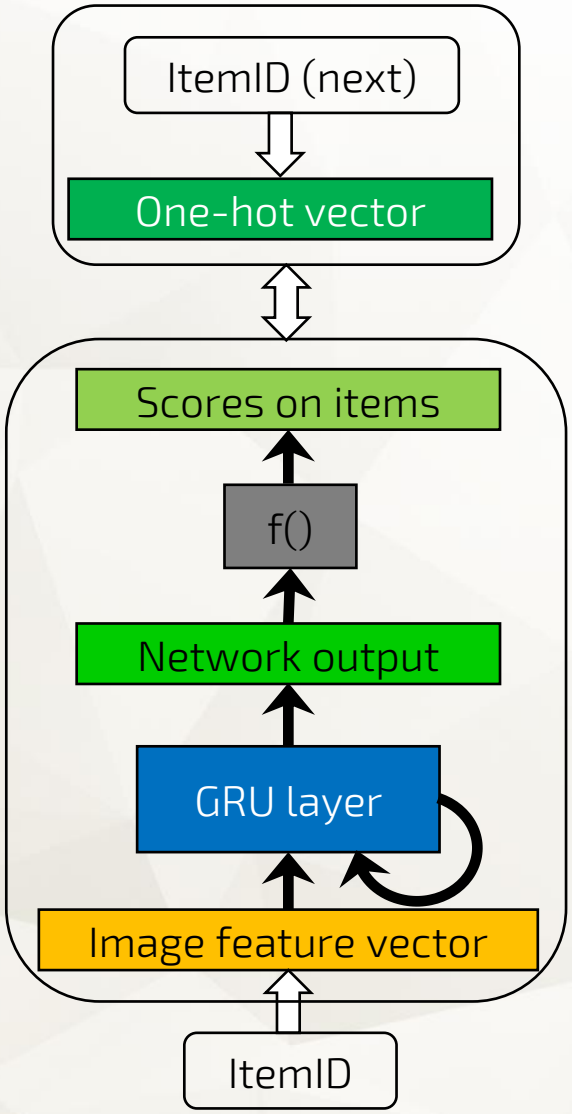
GRAVITY Research & Development

Item features & user decisions

- Clicking on recommendations
- Influencing factors
  - Prior knowledge of the user
  - The stuff they see
    - **Image (thumbnail, product image, etc)**
    - **Text (title, short description, etc)**
    - Price
    - Item properties
    - ...
- Feature extraction from images and text

GRAVITY
Research & Development

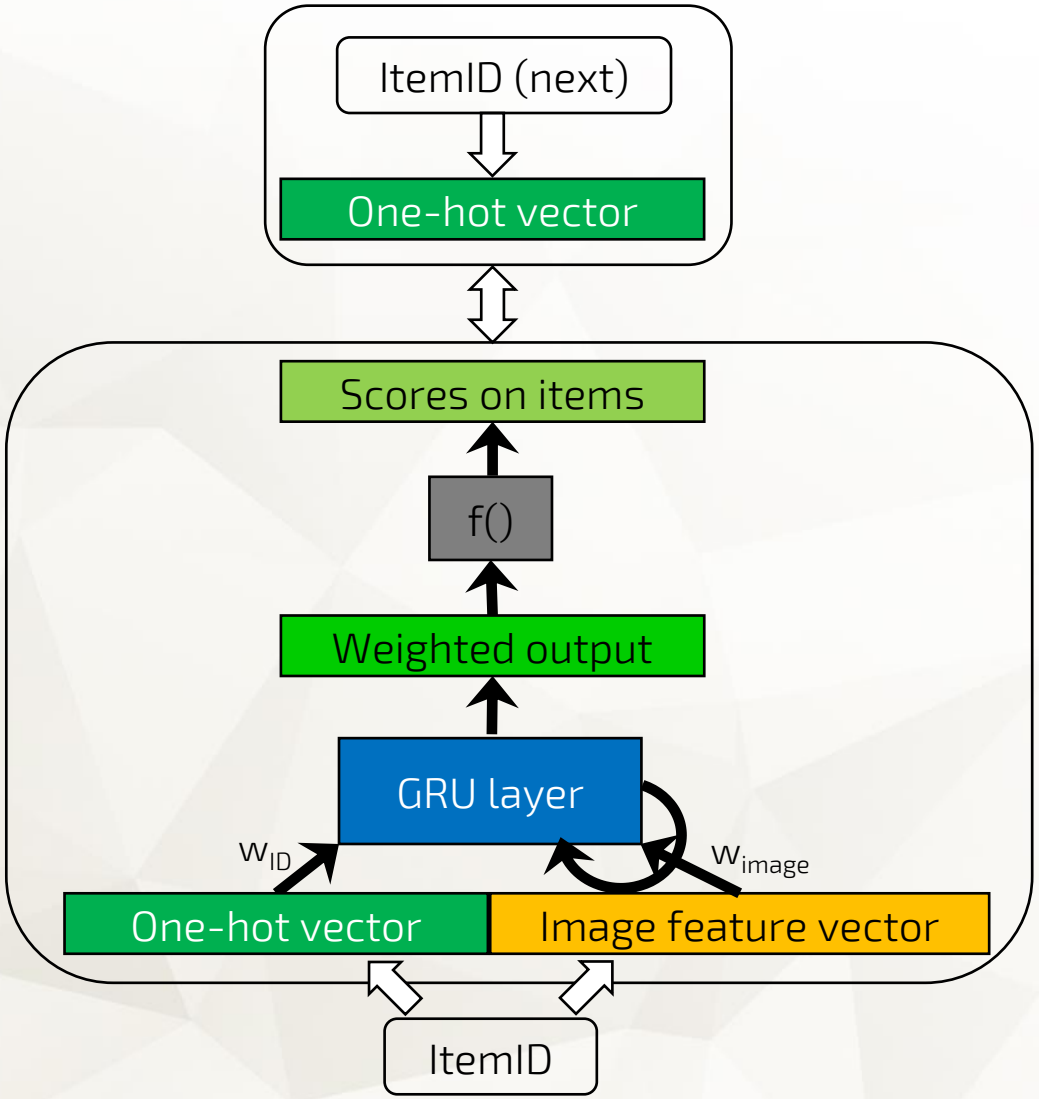
Naive solution 1: Train on the features



- Item features on the input
- Bad offline accuracy
- Recommendations make sense
  - but somewhat general

GRAVITY
Research & Development

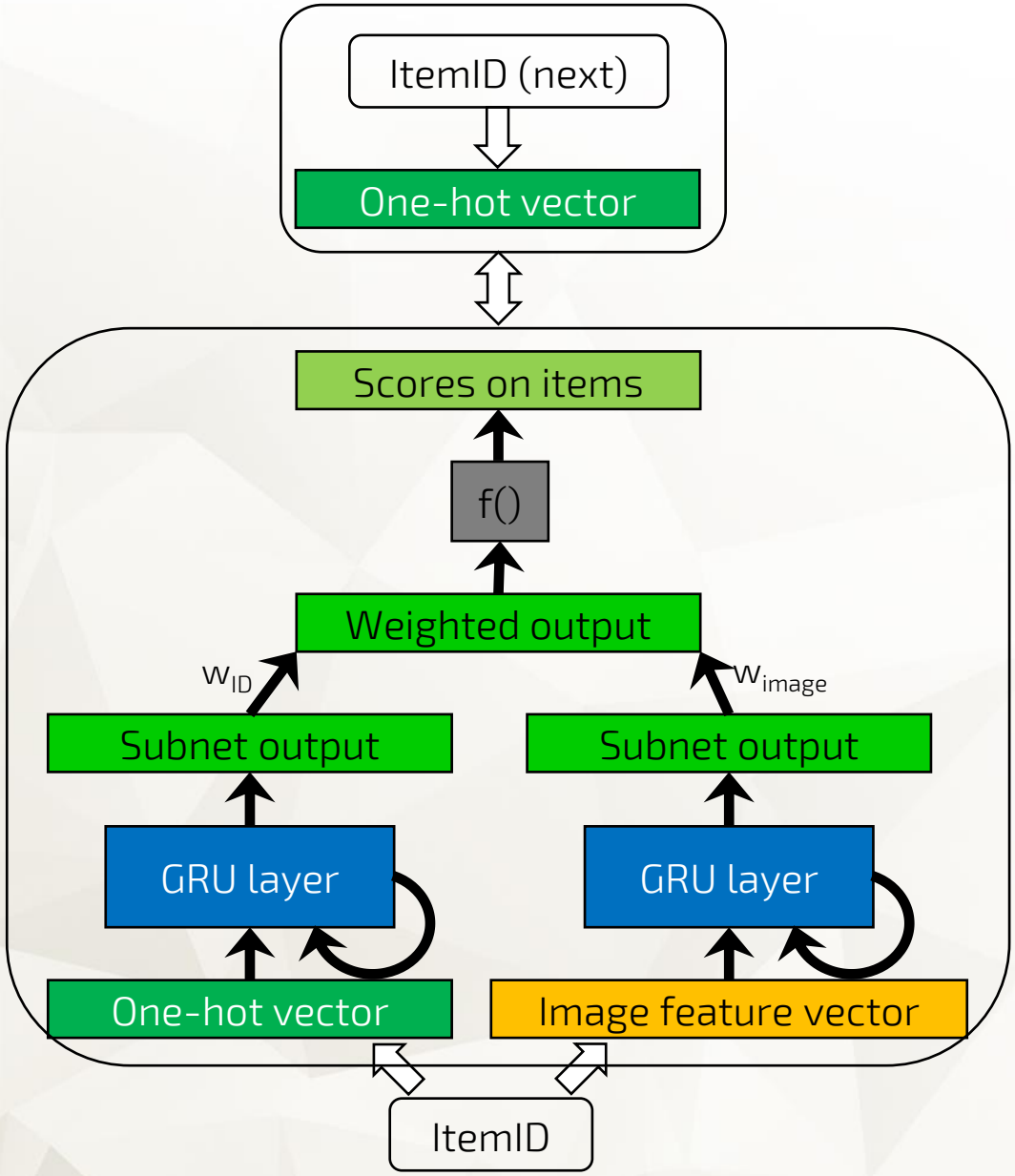
Naive solution 2: Concat. session & feature info



- Concatenate one-hot vector of the ID and the feature vector of the item
- Recommendation accuracy similar to the ID only network
- ID input is initially stronger signal on the next item
  - Dominates the image input
  - Network learns to focus on weights of the ID

GRAVITY
Research & Development

Parallel RNN



- Separate RNNs for each input type
  - ID
  - Image feature vector
  - …
- Subnets model the sessions separately
- Concatenate hidden layers
  - Optional scaling / weighting
- Output computed from the concat. hidden state

GRAVITY
Research & Development

Training

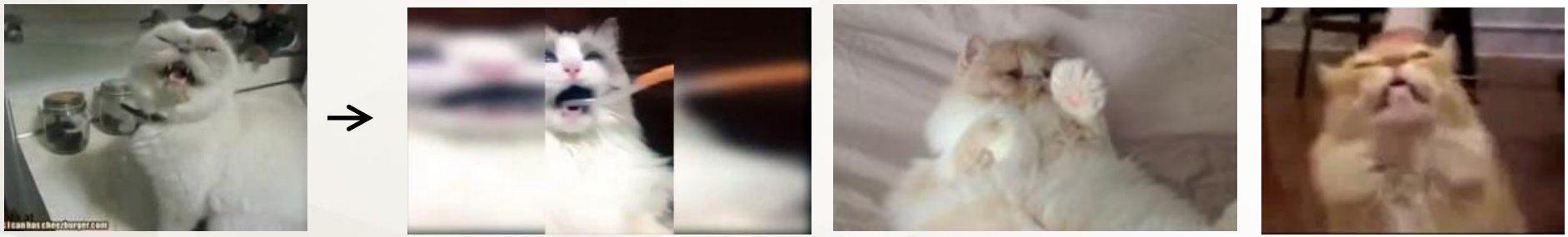
- Backpropagate through the network in one go
- Accuracy is slightly better than for the ID only network
- Subnets learn similar aspects of the sessions
  - Some of the model capacity is wasted

Alternative training methods

- Force the networks to learn different aspects
  - Complement each other
- Inspired by ALS and ensemble learning
- Train one subnet at a time and fix the others
  - Alternate between them
- Depending on the frequency of alternation
  - Residual training (train fully)
  - Alternating training (per epoch)
  - Interleaving training (per minibatch)

Experiments on video thumbnails

- Online video site
  - 712,824 items
  - Training: 17,419,964 sessions; 69,312,698 events
  - Test: 216,725 sessions; 921,202 events
    - Target score compared to only the 50K most popular items
- Feature extraction
  - GoogLeNet (Caffe) pretrained on ImageNet dataset



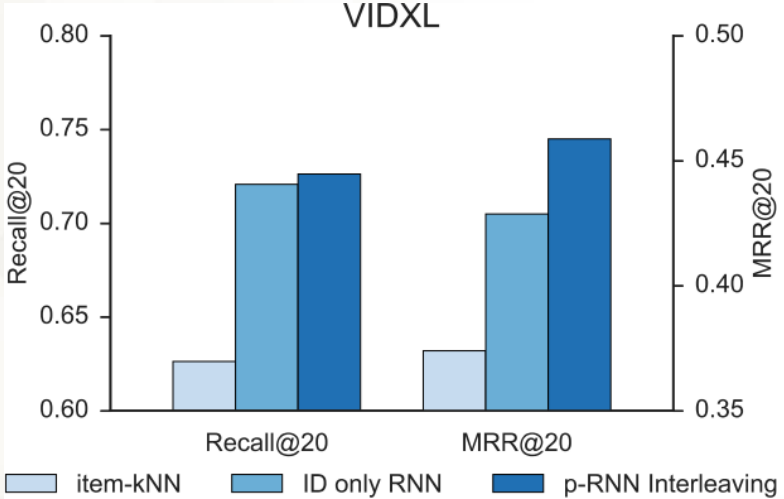
  - Features: values of the last avg. pooling layer
    - Dense vector of 1024 values
    - Length normalized to 1

GRAVITY
Research & Development

Results on video thumbnails

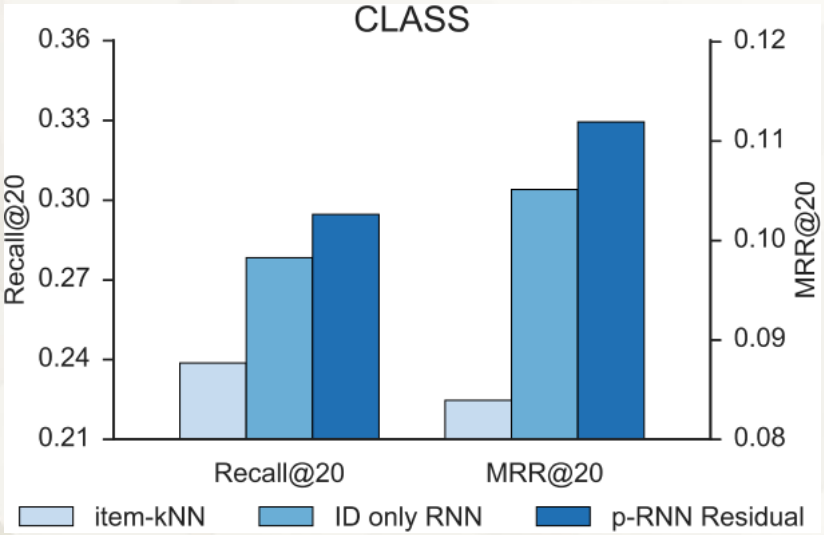
- Low number of hidden units
  - MRR improvements
    - +18.72% (vs item-kNN)
    - +15.41% (vs ID only, 100)
    - +14.40% (vs ID only, 200)
- High number of hidden units
  - 500+500 for p-RNN
  - 1000 for others
  - Diminishing return for increasing the number of epochs or the hidden units
  - +5-7% in MRR

Method	#Hidden units	Recall@20	MRR@20
Item-kNN	-	0.6263	0.3740
ID only	100	0.6831	0.3847
ID only	200	0.6963	0.3881
Feature only	100	0.5367	0.3065
Concat	100	0.6766	0.3850
P-RNN (naive)	100+100	0.6765	0.4014
P-RNN (res)	100+100	0.7028	**0.4440**
P-RNN (alt)	100+100	0.6874	0.4331
P-RNN (int)	100+100	**0.7040**	0.4361



Experiments on text

- Classified ad site
  - 339,055 items
  - Train: 1,173,094 sessions; 9,011,321 events
  - Test: 35,741 sessions; 254,857 events
    - Target score compared with all items
  - Multilanguage user generated titles and item descriptions
- Feature extraction
  - Concatenated title and description
  - Bi-grams with TF-IDF weighting
    - Sparse vectors of 1,099,425 length; 5.44 avg. non-zeroes



- Experiments with high number of hidden units
- +3.4% in recall
- +5.37% in MRR

GRAVITY
Research & Development

Summary

- Incorporated image and text features into session-based recommendations using RNNs
- Naive inclusion is not effective
- P-RNN architectures
- Training methods for P-RNNs

GRAVITY
Research & Development

Future work

- Revisit feature extraction
- Try representations of the whole video
- Use multiple aspects for items

GRAVITY
Research & Development

Thanks!

Q&A

GRAVITY
Research & Development