Location Aware Sequential Recommendation Using Hybrid Recurrent Neural Networks

Dimitris Michailidis

26th June 2019

Supervisors: Dmitri Jarnikov (Naspers)
Pengjie Ren (ILPS, University of Amsterdam)

University of Amsterdam

Sequential Recommendation

Given a sequence of actions:

$$[x_1, x_2, .., x_\tau, .., x_{t-1}]$$

Find a model F that predicts:

$$y = F([x_1, x_2, .., x_{t-1}]) = x_t$$

Sequential Recommendation

Given a sequence of actions:

$$[x_1, x_2, .., x_\tau, .., x_{t-1}]$$

Find a model F that predicts:

$$y = F([x_1, x_2, .., x_{t-1}]) = x_t$$

Sequential Recommendation

Given a sequence of actions:

$$[x_1, x_2, .., x_\tau, .., x_{t-1}]$$

Find a model F that predicts:

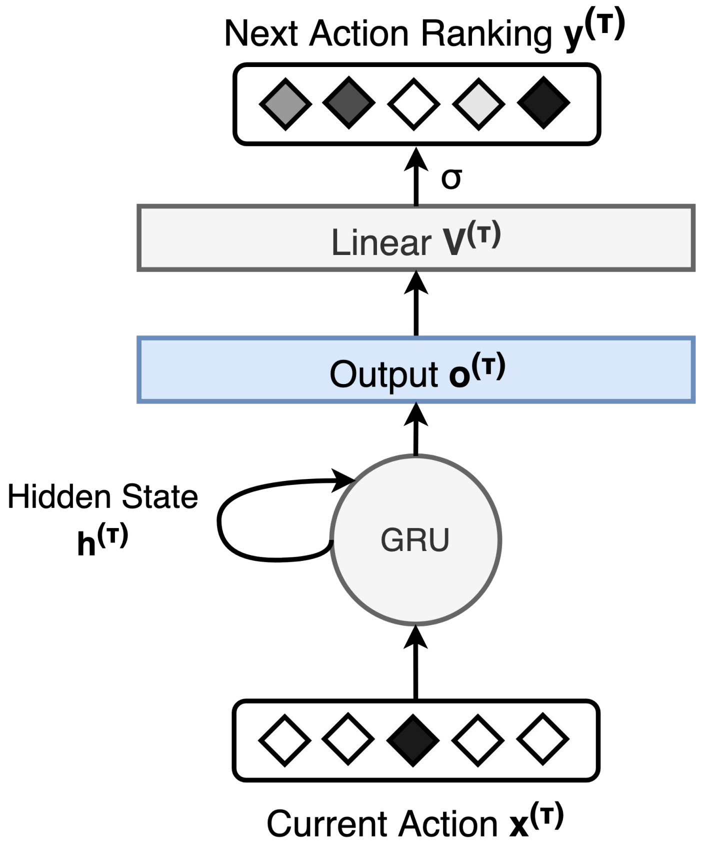
$$y = F([x_1, x_2, .., x_{t-1}]) = x_t$$

Latest Advancements

Intuition

- **RNNs** have shown **great improvements** in Language Modeling.
- We can **represent actions** the **same way as words** in a sentence.
- Use the **hidden state** to hold information about the sequence.

Great improvements, inspired all follow-up works.



(Hidasi et al. ICLR 2016)

Latest Advancements

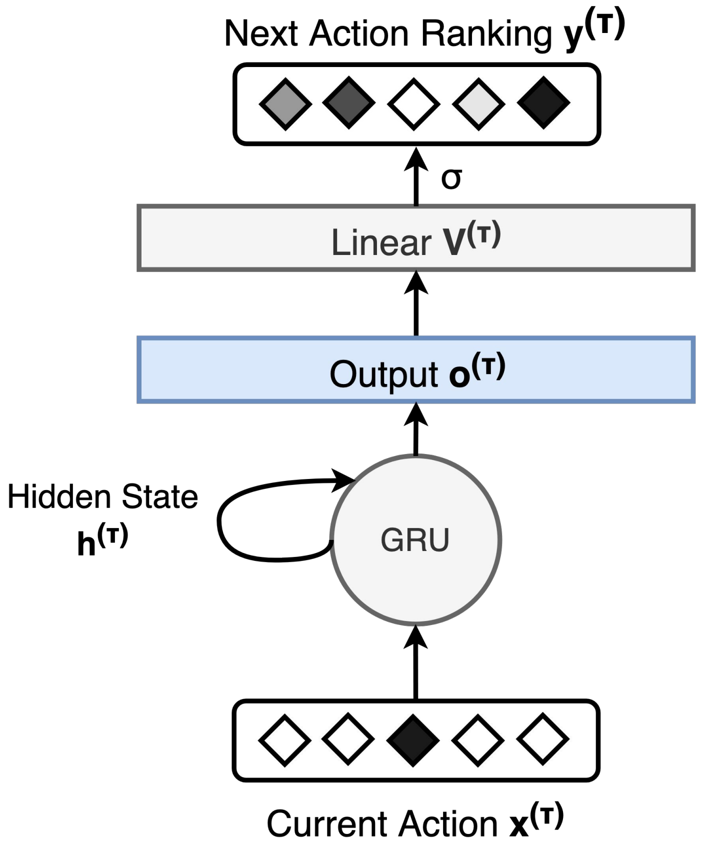
Intuition

- **RNNs** have shown **great improvements** in Language Modeling.
- We can **represent actions** the **same way as words** in a sentence.
- Use the **hidden state** to hold information about the sequence.

Great improvements, inspired all follow-up works.

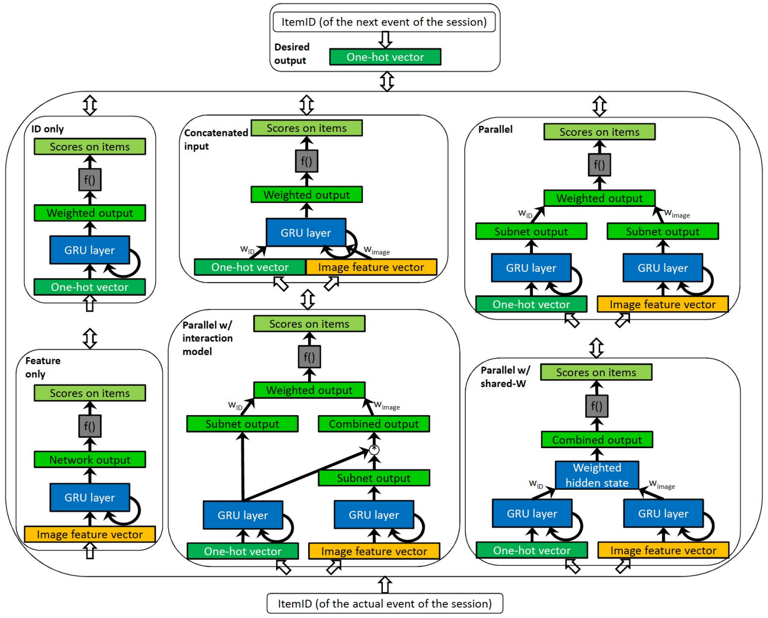
Fallbacks

- Only depends on clicked items, ignores **context**.
    - **Visual:** thumbnail, title, description
    - **Scope:** home-page, side-page, checkout
    - **Evolution:** time

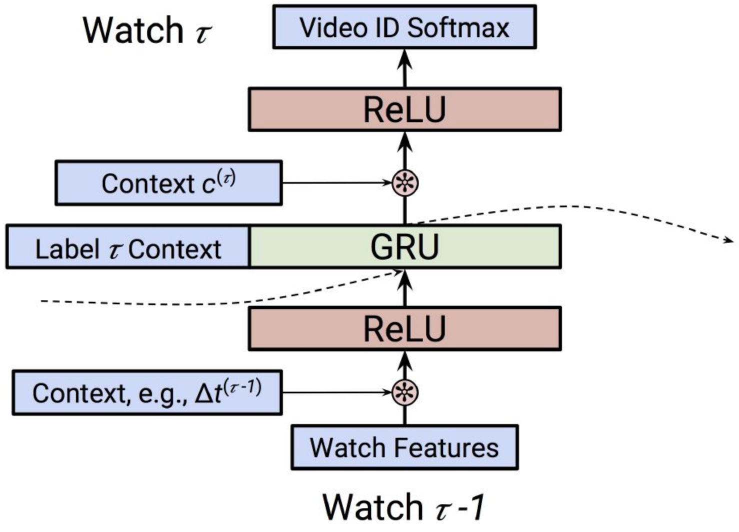


(Hidasi et al. ICLR 2016)

Contextual Recurrent Recommendation



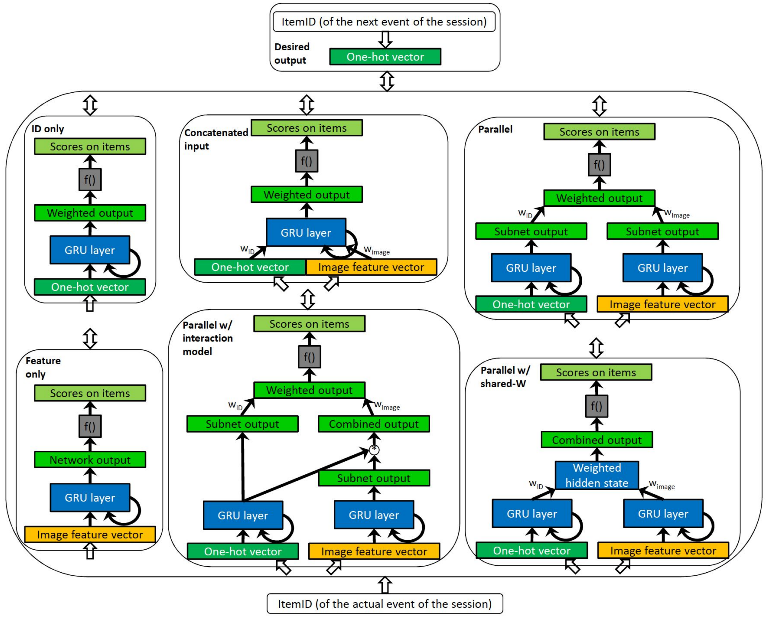
(Hidasi et al. RecSys 2016)



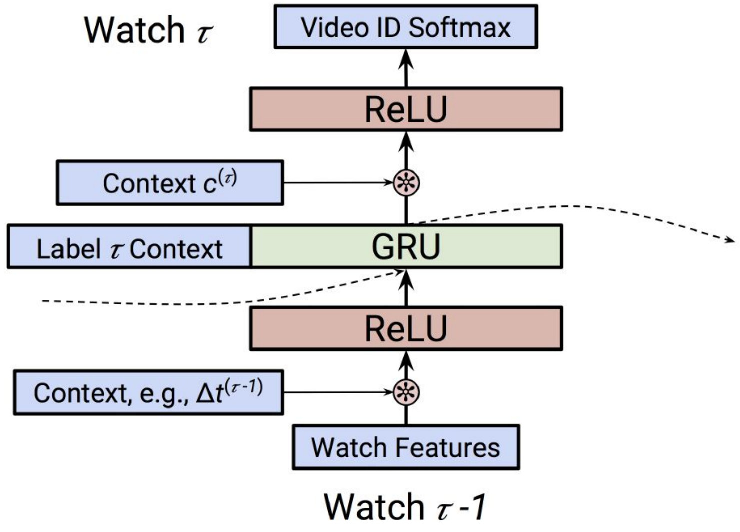
(Beutel et al. WSDM 2018)

Contextual Recurrent Recommendation

Retail Domain, **Text** and **Image** context



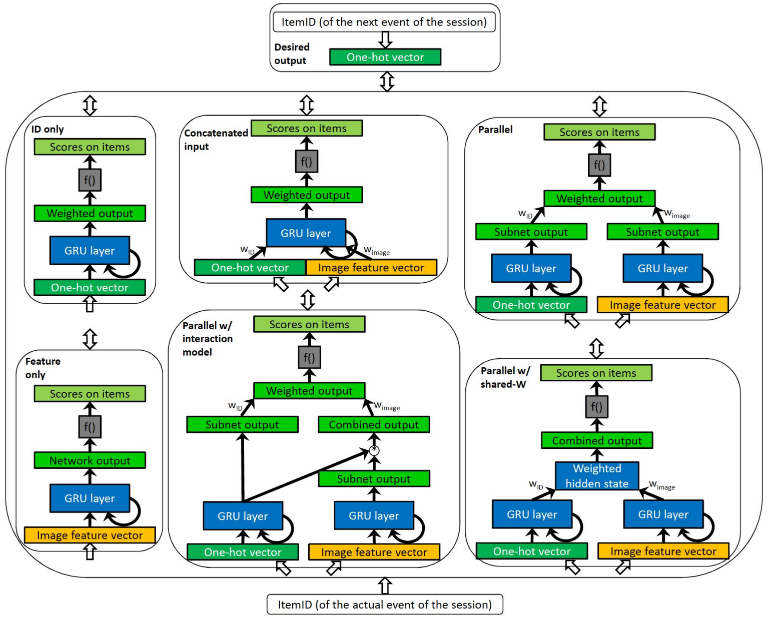
(Hidasi et al. RecSys 2016)



(Beutel et al. WSDM 2018)

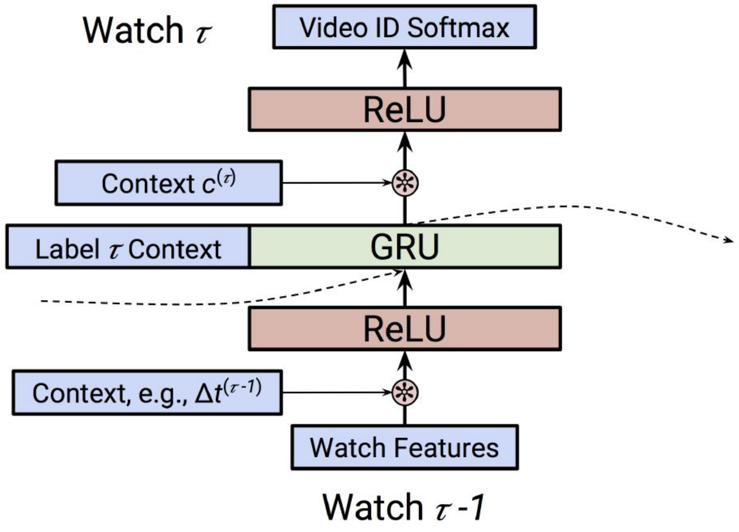
Contextual Recurrent Recommendation

Retail Domain, **Text** and **Image** context



(Hidasi et al. RecSys 2016)

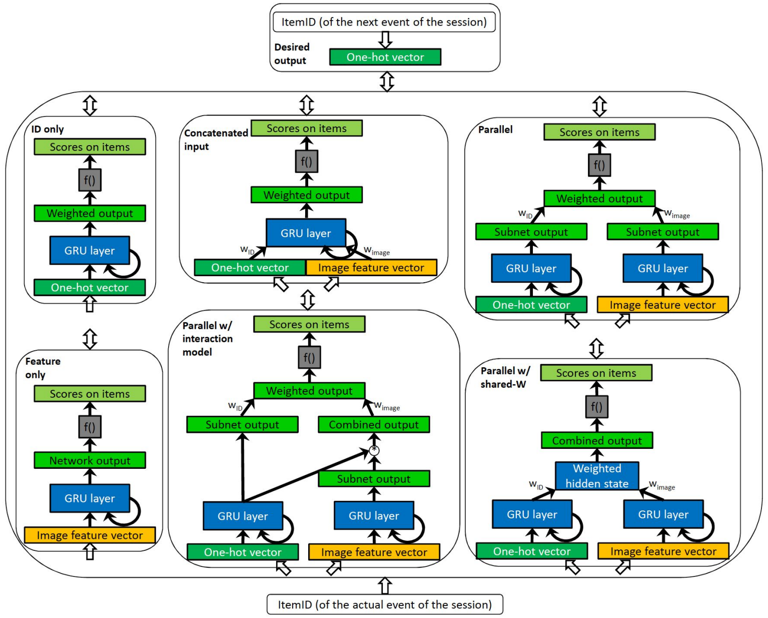
Video Domain, **Time** context



(Beutel et al. WSDM 2018)

Contextual Recurrent Recommendation

Retail Domain, **Text** and **Image** context

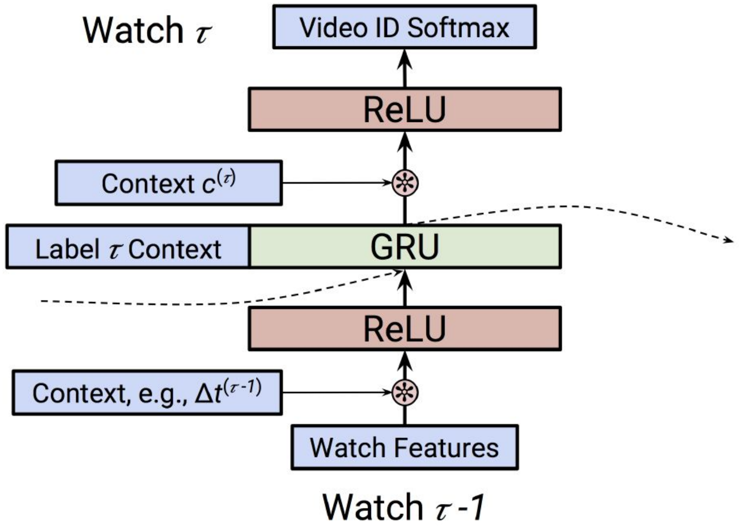


(Hidasi et al. RecSys 2016)

Our Approach:

- **Peer-to-Peer marketplace** domain
- **Location** Context
- **Combine** ideas from both.

Video Domain, **Time** context



(Beutel et al. WSDM 2018)

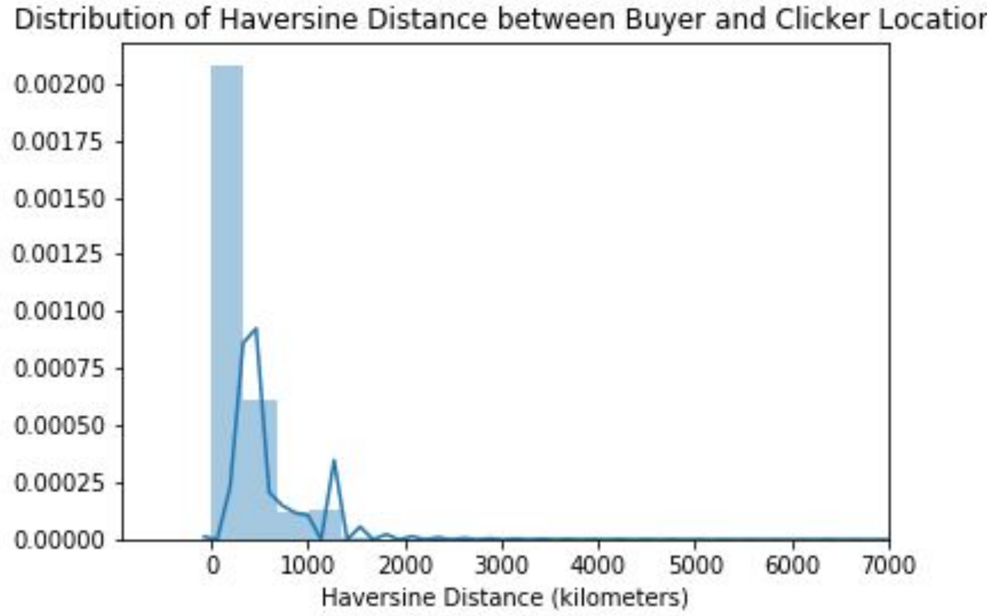
Why location?

Motivation

- Current research focuses on **buyer-centric** context.
- What about **buyer-seller** dependencies?

Distance alone might not be enough

- Travel time
- Accessibility
- Means of Transport
- Neighborhood security



Data from OLX South Africa

Representing Location

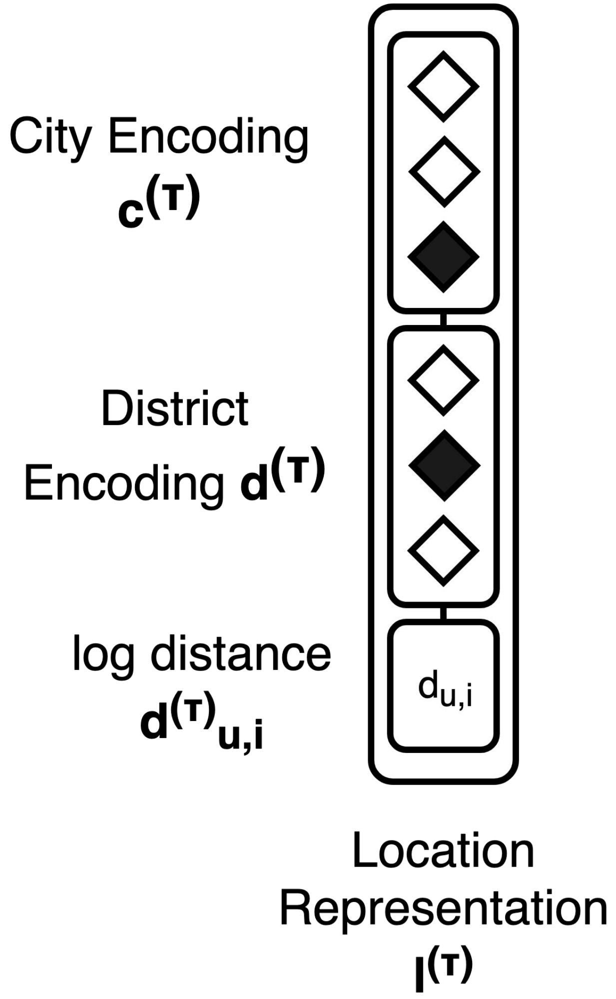
For each user (buyer/seller) we have the following information:

- Geolocation approximation
  - Country
  - City
  - District

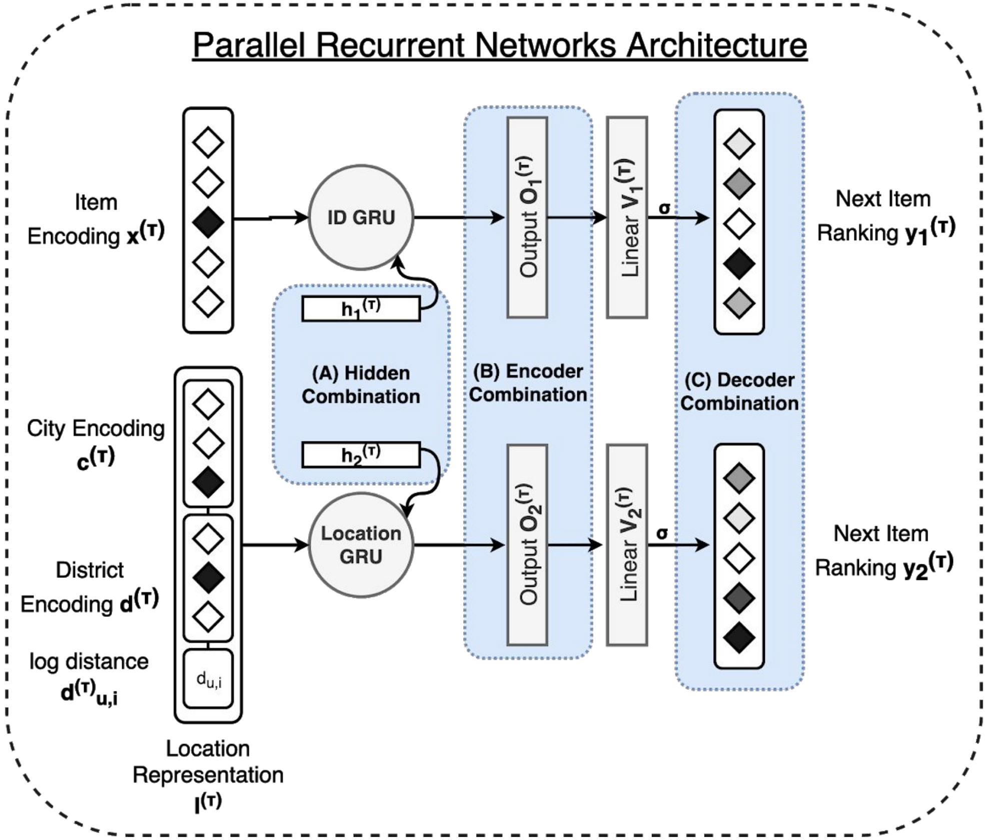
Embed the city and district using **1-hot-encoding**.

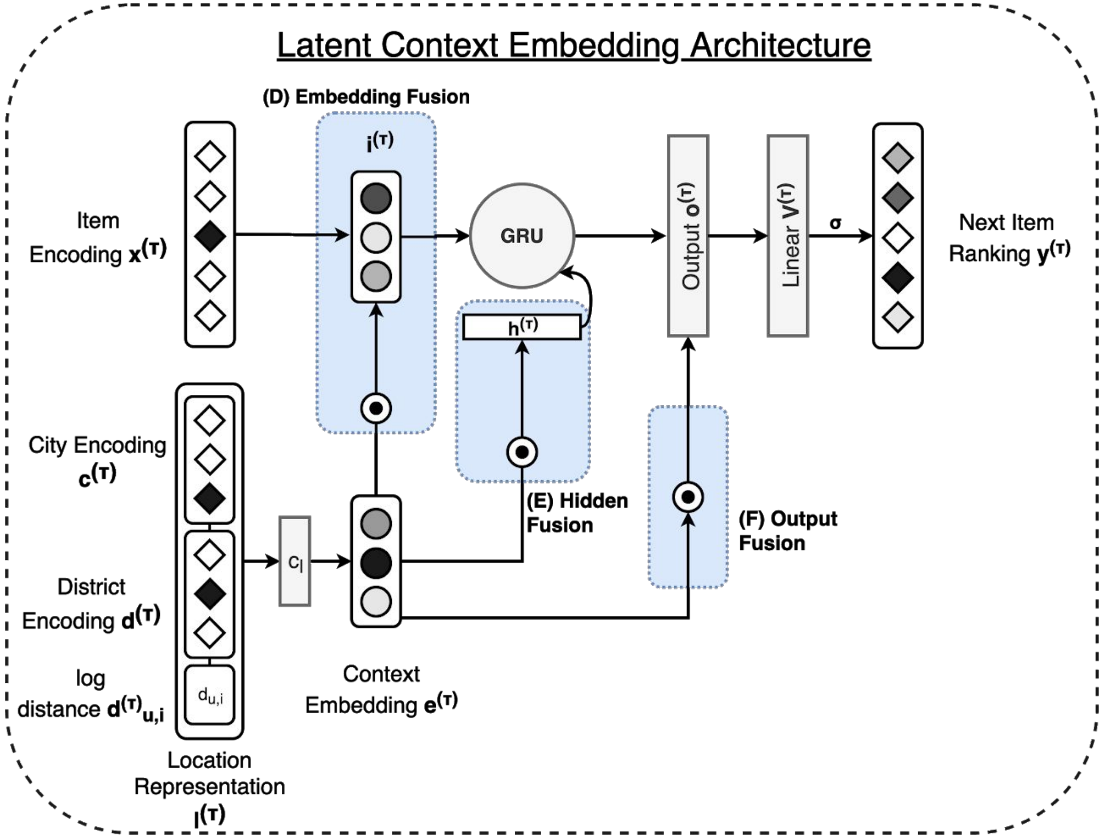
Calculate the **log distance** between the users.

Concatenate them.



Proposed Architectures





Parallel Recurrent Architecture

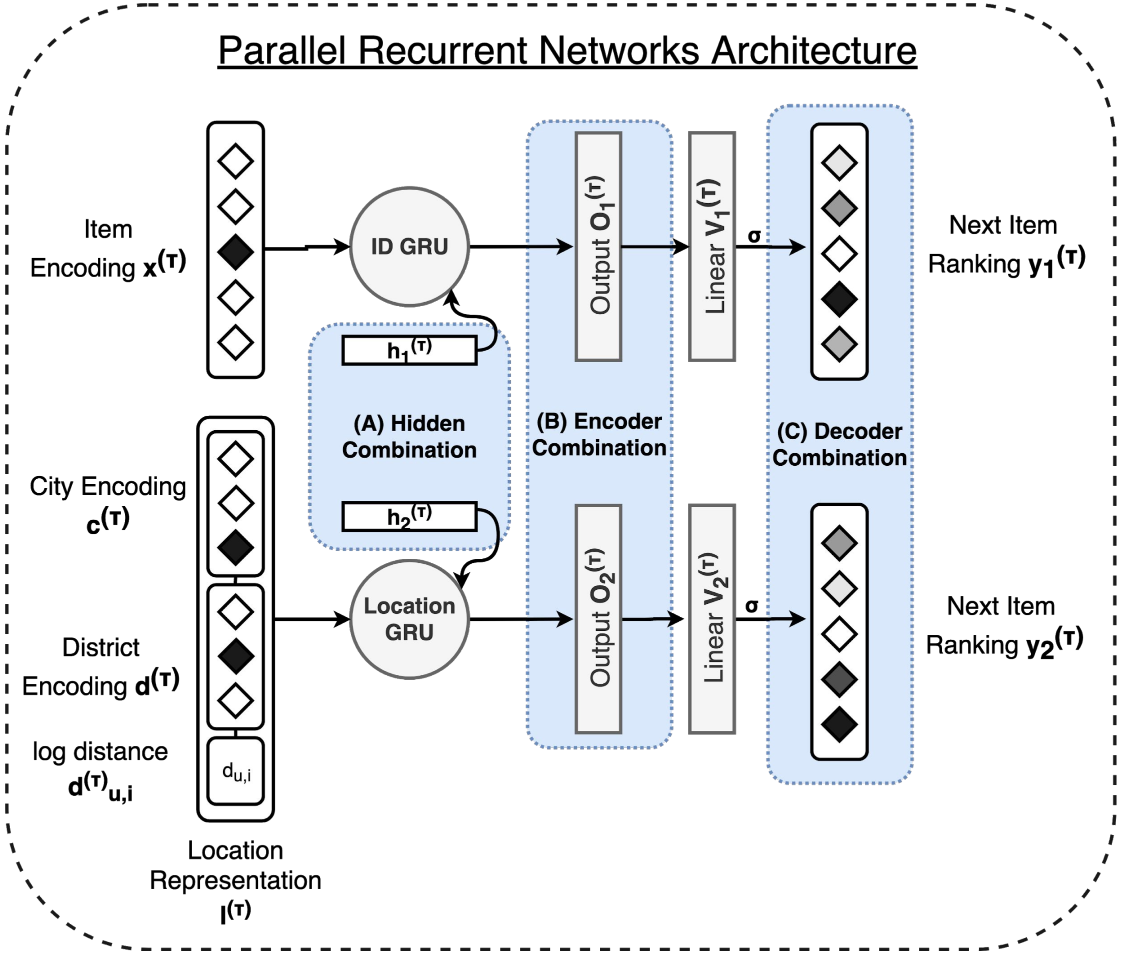
Train **parallel recurrent** models that learn from:

1) the clicked **items** and

2) the **location representation**

Proposed combinations:

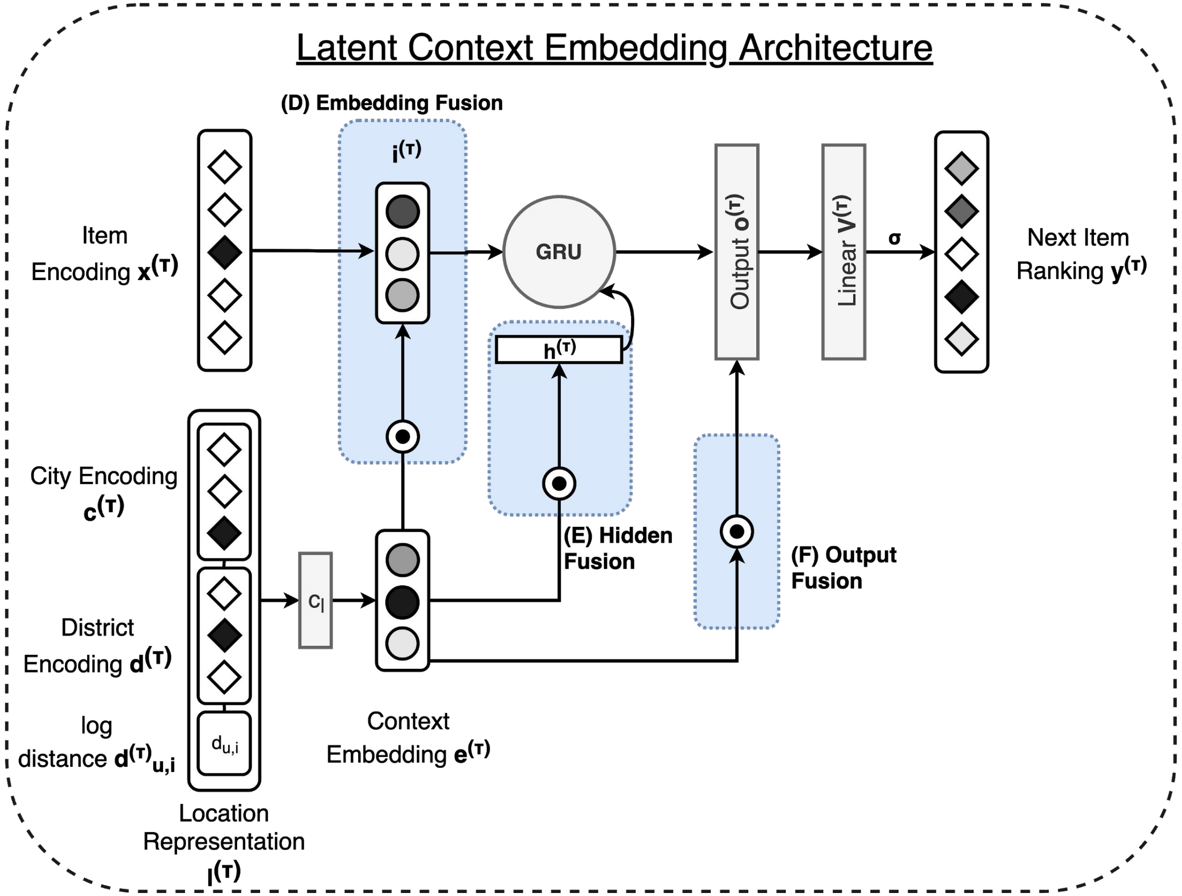
A) Hidden state level
B) Encoder level (GRU output)
C) Decoder level (before final activation)



Latent Context Architecture

Location context is modeled with a **latent representation**.

Contextualize the base GRU by fusing the location context in different stages:



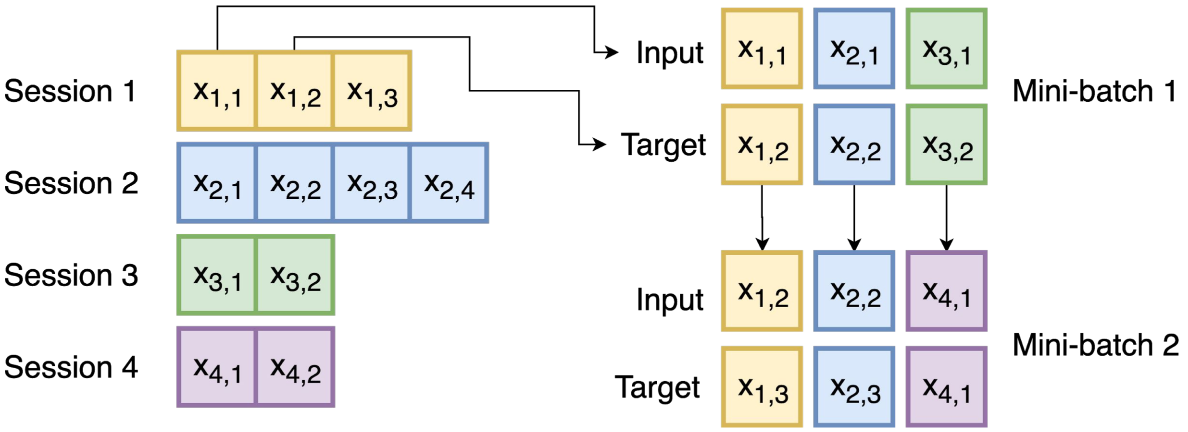
D) Before the Recurrent Phase (Embedding Fusion)

E) During the Recurrent Phase (Hidden Fusion)

F) After the Recurrent Phase (Decoder Fusion)

Training

Session-parallel mini-batches, based on (Hidasi et al. 2015)



TOP1 **Pairwise-Ranking** Loss Function (Hidasi et al. 2016)

$$L_s = \frac{1}{N_s} \sum_{j=1}^{N_s} \sigma(r_{s,j} - r_{s,i}) + \sigma({r^2}_{s,j})$$

Experimental Setup

Research Questions

RQ1. Do the proposed **parallel recurrent architectures** outperform the baselines?

RQ2. Does modeling location information with a **latent representation** perform better than the parallel modeling?

RQ3. What is the **best** way to represent location? Does **distance** alone **suffice**?

RQ4. What is the best way to **combine the independent parts** in the recurrent approaches?

Experimental Setup

Dataset

- OLX Impressions on Vehicles, gathered for a time-span of a week.
  - Train on the first six days, test on the final.
- ~ 250.000 click actions
- ~ 13.000 unique items
- ~ 4.5 average session length
- ~ 261km average distance between buyer and seller

Experimental Setup

Baselines

Non-RNN:

- **POP/SPOP**: Most popular items overall/in session [14, 16].
- **LocationPOP**: Most popular items nearby. Devised by us.
- **ItemKNN**: Item-to-item CF, strongest non-RNN baseline [9, 22]

RNN-Baselines:

- **GRU4Rec**: Simple GRU that learns from clicked items [16]
- **LocGRU**: Uses only location to predict next items. Devised by us.

Experimental Setup

Evaluation

- **Mean Recall@k**: Percentage of cases the next item appears on the top@k list.
    - Highly correlated with click-through rate (CTR) [16, 23].
- **MRR@k** : 1/rank of the next item
    - The most improved metric compared to the baseline in similar works [17, 28].

Evaluated at k = [5, 10, 20]

Results (Non-RNN)

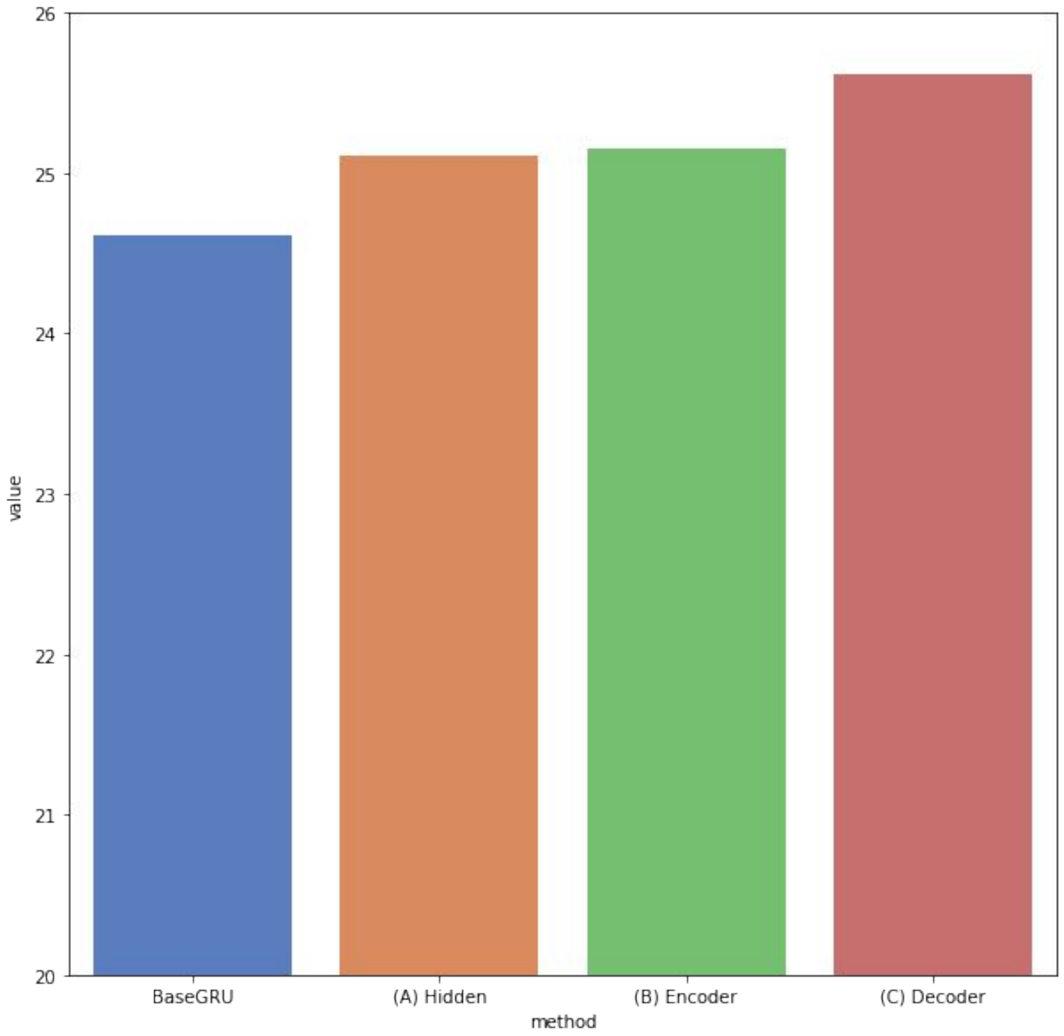
(Results in %)

Model	Recall@20	MRR@20
POP	1.03	0.24
SPOP	0.83	0.11
LocationPOP	2.03	0.44
Item KNN	15.15	4.56

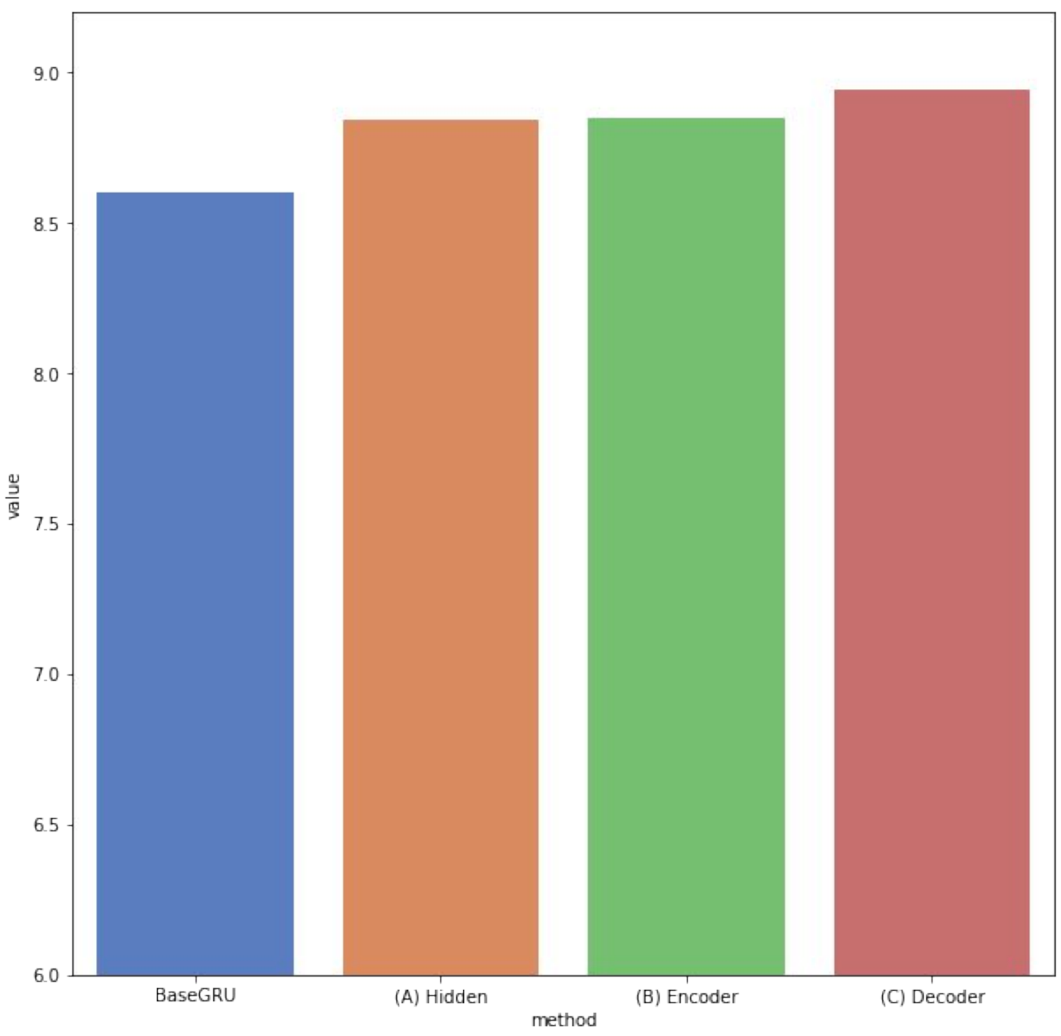
All proposed models outperform non-rnn baselines by a large margin.

Results (RQ1)

Recall@20



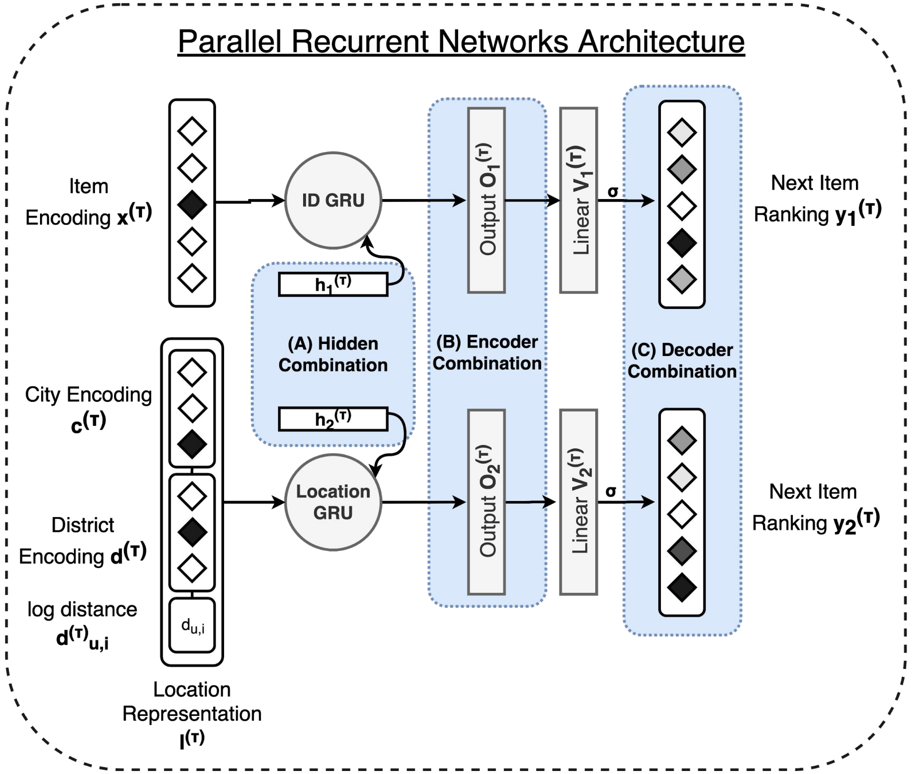
MRR@20



Results (RQ1)

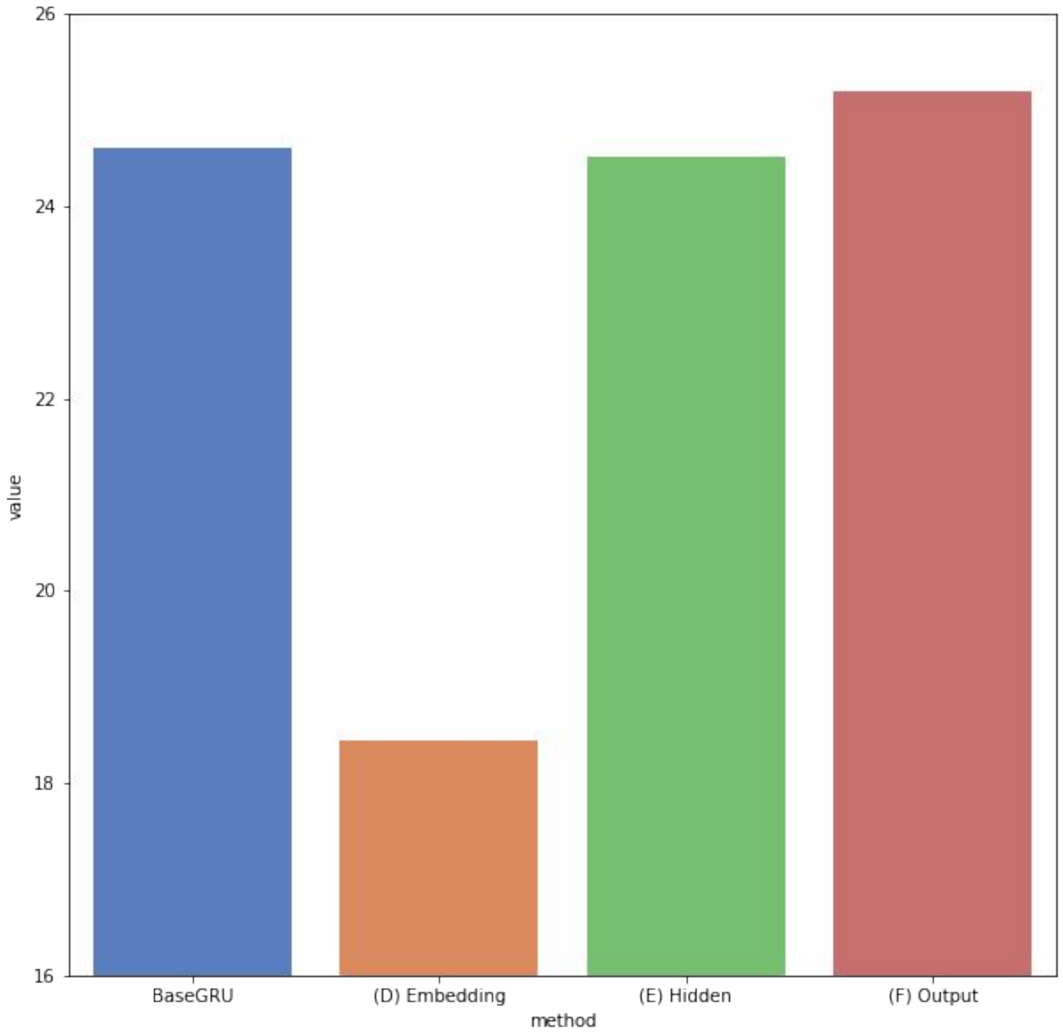
Overview

- **(C) Decoder Combination** is the best performing model overall.
    - Improves **Recall** by **3.3%** on average.
    - Improves **MRR** by **4.1%** on average.
- It also outperforms the other parallel architectures.
- Why?
    - The recurrent units **learn from different data**.
    - But the end **goal is the same**.

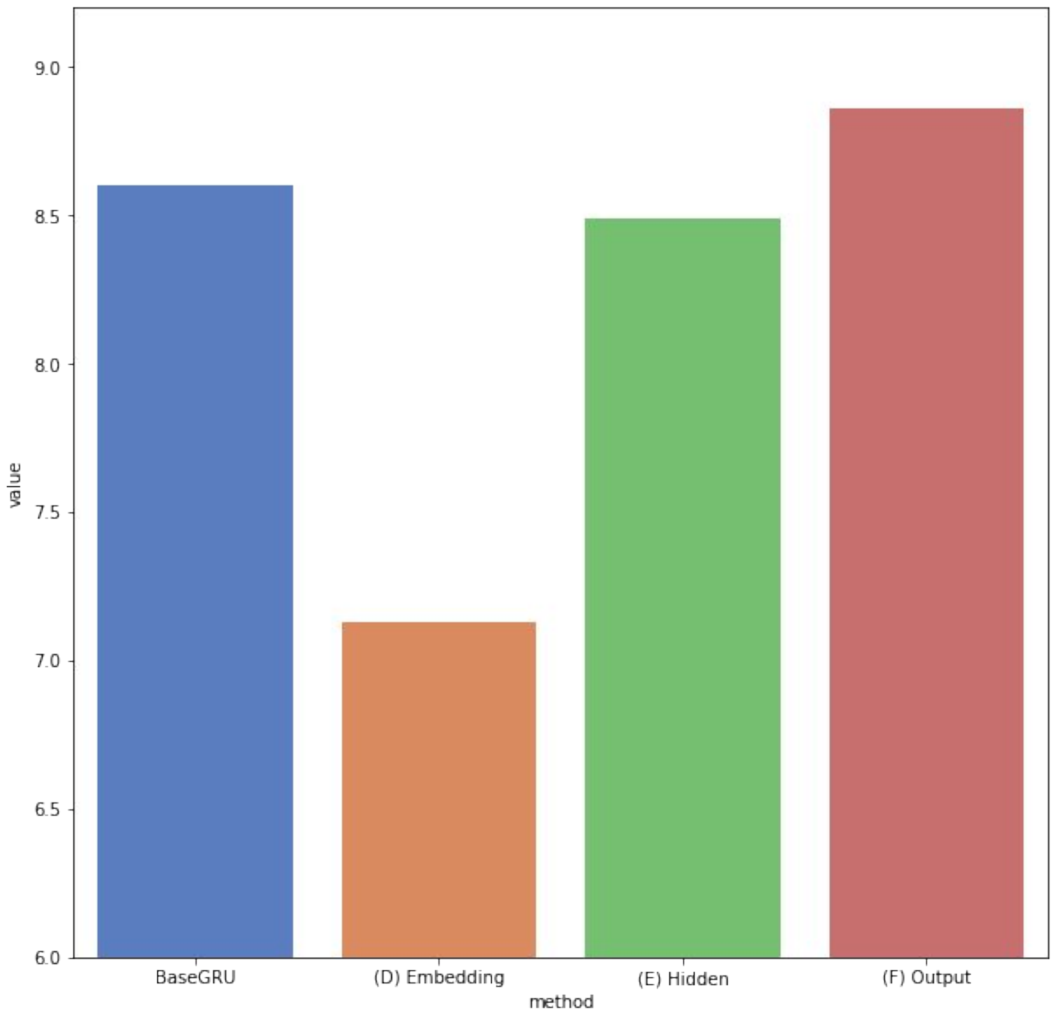


Results (RQ2)

Recall@20



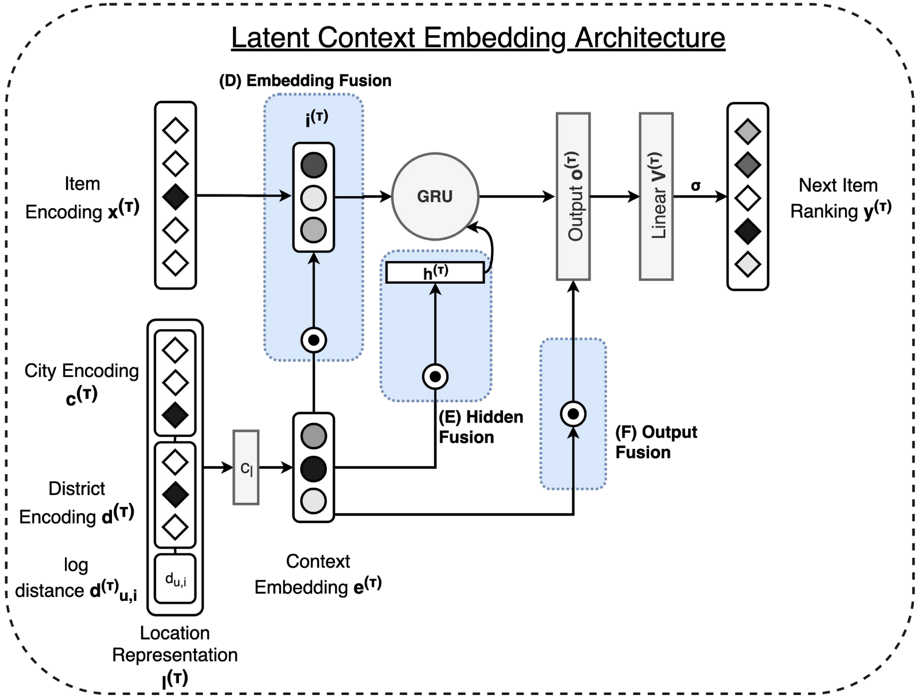
MRR@20



Results (RQ2)

Overview

- Contextualizing the event embedding with location in the input level is **not sufficient,** contrary to [21].
- **(F) Output Fusion** is the only one that beats the baseline.
    - **2.9%** increase in **Recall**.
    - **3.5%** increase in MRR.



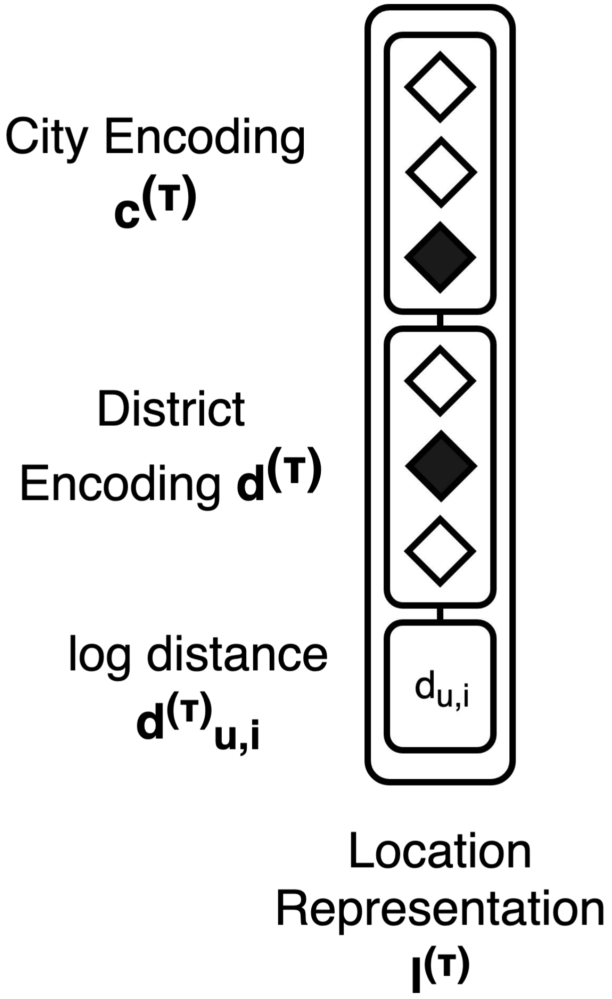
Why?

- The base recurrent unit is already good enough. Additional context should be considered after it.

Analysis (RQ3)

	Parallel Decoder Model					
	Recall			MRR		
Method	@5	@10	@20	@5	@10	@20
City Encoding	13.15	18.80	25.50	7.61	8.35	8.81
District Encoding	13.13	18.81	25.47	7.62	8.37	8.82
Distance	13.26	18.67	25.55	7.54	8.26	8.73
Full Concatenation	**13.30**	**18.92**	**25.62**	**7.73**	**8.48**	**8.94**

	Latent Output Fusion Model					
	Recall			MRR		
Method	@5	@10	@20	@5	@10	@20
City Encoding	13.19	18.59	25.11	7.67	8.38	8.84
District Encoding	13.28	18.66	25.11	7.63	8.34	8.78
Distance	**13.32**	**18.80**	**25.20**	**7.70**	**8.42**	**8.86**
Full Concatenation	**13.32**	18.65	25.18	**7.70**	8.40	8.85



Analysis (RQ4)

Method	Recall			MRR		
	@5	@10	@20	@5	@10	@20
Addition	11.68	15.93	20.78	6.82	7.38	7.72
Multiplication	11.53	16.11	20.79	6.57	7.18	7.50
Weighted Average	**13.22**	**18.92**	**25.62**	**7.73**	**8.48**	**8.94**

Overview

- **Addition:** as a less complex approach ([5, 30]), underperforms compared to the others.
- **Multiplication**: location not enough to apply context to the base network.
- **Weighted Average:** the best approach, results in alignment with [17].

Limitations & Future Work

- Parallel models add additional complexity to the model.
    - Double RNN units means double parameters to learn.
- Models were trained simultaneously, not by using residual or interleaving training.
    - Researches have shown that these can perform even better [17].
- Relatively small dataset, therefore contextual embeddings could be learned better.
- Dataset came from interactions logs which were driven by an already present recommender system.

Thank you!

Questions?

Extra Slides

30

Overall Results

Table 1: Experimental Results on the OLX Vehicle Dataset. Values are in percentages.

	OLX Vehicles Dataset					
	Recall			MRR		
Models	@5	@10	@20	@5	@10	@20
POP	0.26	0.49	1.03	0.17	0.20	0.24
S-POP	0.14	0.35	0.83	0.05	0.08	0.11
LOC-POP	0.53	1.02	2.03	0.29	0.36	0.44
Item-KNN	6.04	10.20	15.15	3.56	4.20	4.56
LocGRU	5.20	7.89	11.46	2.89	3.24	3.48
BaseGRU	12.90	18.25	24.61	7.45	8.10	8.60
(A) Parallel Hidden	13.02	18.58	25.11	7.51	8.31	8.84
(B) Parallel Encoder	13.12	18.73	25.15	7.66	8.41	8.85
(C) Parallel Decoder	13.30	**18.92**	**25.62**	**7.73**	**8.48**	**8.94**
(D) Embedding Fusion	10.83	14.70	18.44	6.35	6.87	7.13
(E) Hidden Fusion	12.76	18.26	24.52	7.33	8.06	8.49
(F) Output Fusion	**13.32**	18.80	25.20	7.70	8.42	8.86

Bold face indicates best result in the corresponding metric.

Overall Results

Table 2: Effect of different Location Representations in the Parallel Decoder architecture.

	Parallel Decoder Model					
	Recall			MRR		
Method	@5	@10	@20	@5	@10	@20
City Encoding	13.15	18.80	25.50	7.61	8.35	8.81
District Encoding	13.13	18.81	25.47	7.62	8.37	8.82
Distance	13.26	18.67	25.55	7.54	8.26	8.73
Full Concatenation	**13.30**	**18.92**	**25.62**	**7.73**	**8.48**	**8.94**

Table 3: Effect of different Location Representations in the Latent Output Fusion architecture.

	Latent Output Fusion Model					
	Recall			MRR		
Method	@5	@10	@20	@5	@10	@20
City Encoding	13.19	18.59	25.11	7.67	8.38	8.84
District Encoding	13.28	18.66	25.11	7.63	8.34	8.78
Distance	**13.32**	**18.80**	**25.20**	**7.70**	**8.42**	**8.86**
Full Concatenation	**13.32**	18.65	25.18	**7.70**	8.40	8.85

Overall Results

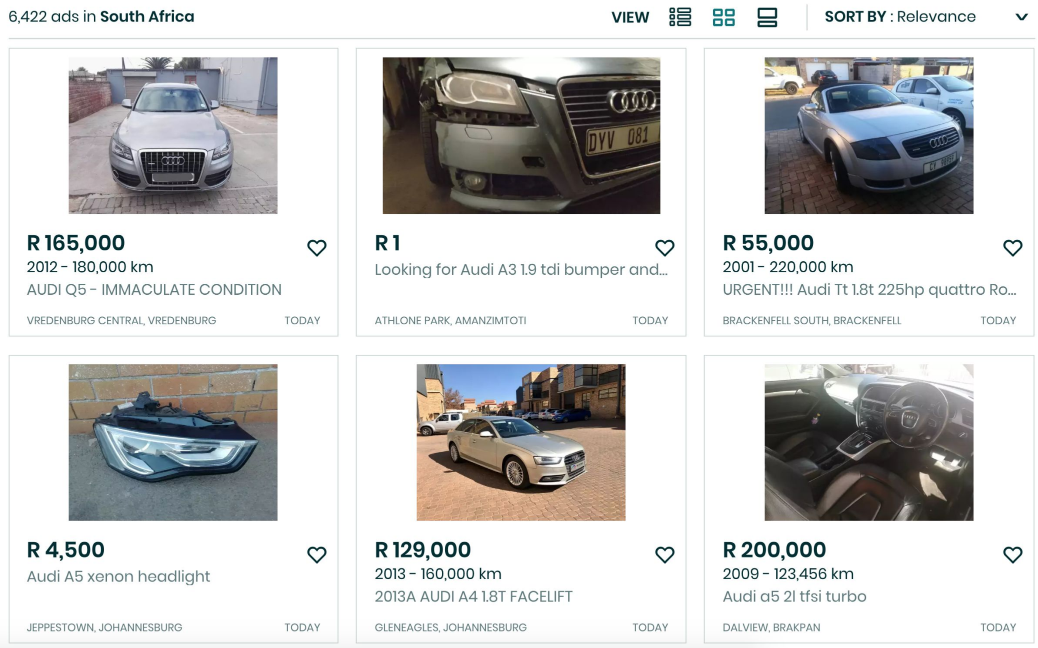
Table 4: Effect of different output combination strategies in the Parallel Recurrent architecture.

	Recall			MRR		
Method	@5	@10	@20	@5	@10	@20
Addition	11.68	15.93	20.78	6.82	7.38	7.72
Multiplication	11.53	16.11	20.79	6.57	7.18	7.50
Weighted Average	**13.22**	**18.92**	**25.62**	**7.73**	**8.48**	**8.94**

Contextual Recommendation

Many factors contribute to a user's interactions:

- Attractiveness of listing: **price**, **image**, **title**
- Current scope: **home-page, side-page, checkout-page**
- Evolution of preference: **time, device**



https://www.olx.co.za