Deep Learning-based Short Video Recommendation and Prefetching for Mobile Commuting Users

Qian Li[1], **Yuan Zhang**[1], Hong Huang[2], Jinyao Yan[1]
1 Communication University of China
2 Huazhong University of Science and Technology

— 2019.8 —

Short Video Applications

- What is short video application?
  - ✓User-generated short video clips
  - ✓Usually <1min
  - ✓Vlog, advertising, self-media, etc.
  - ✓Can only be watched online
- Rapid growth of short video applications[1]
  - ✓Number of short video users is 648 million in China by the end of 2018
  - ✓Number of daily active users is growing at a speed of > 800% each year since 2016

1 The 43rd statistical report on the development of Internet in China.2018

Commuting Scenario

Watching short videos using public transportation (bus, subway, etc.) during way to work

In a survey[1] with 190 respondents, over 46% reported watching short videos on their daily commuting time by public transport service

High moving speed, unstable network connection[2]

Over 71% out of the 46% respondents reported having disconnections when watching short videos on public transportation

Regular, predictable

1 https://www.wjx.cn/m/37307561.aspx
2 H. Deng, et. al., Mobility Support in Cellular Networks: A Measurement Study on Its Configurations and Implications. IMC '18

Problem and Objective

User QoE

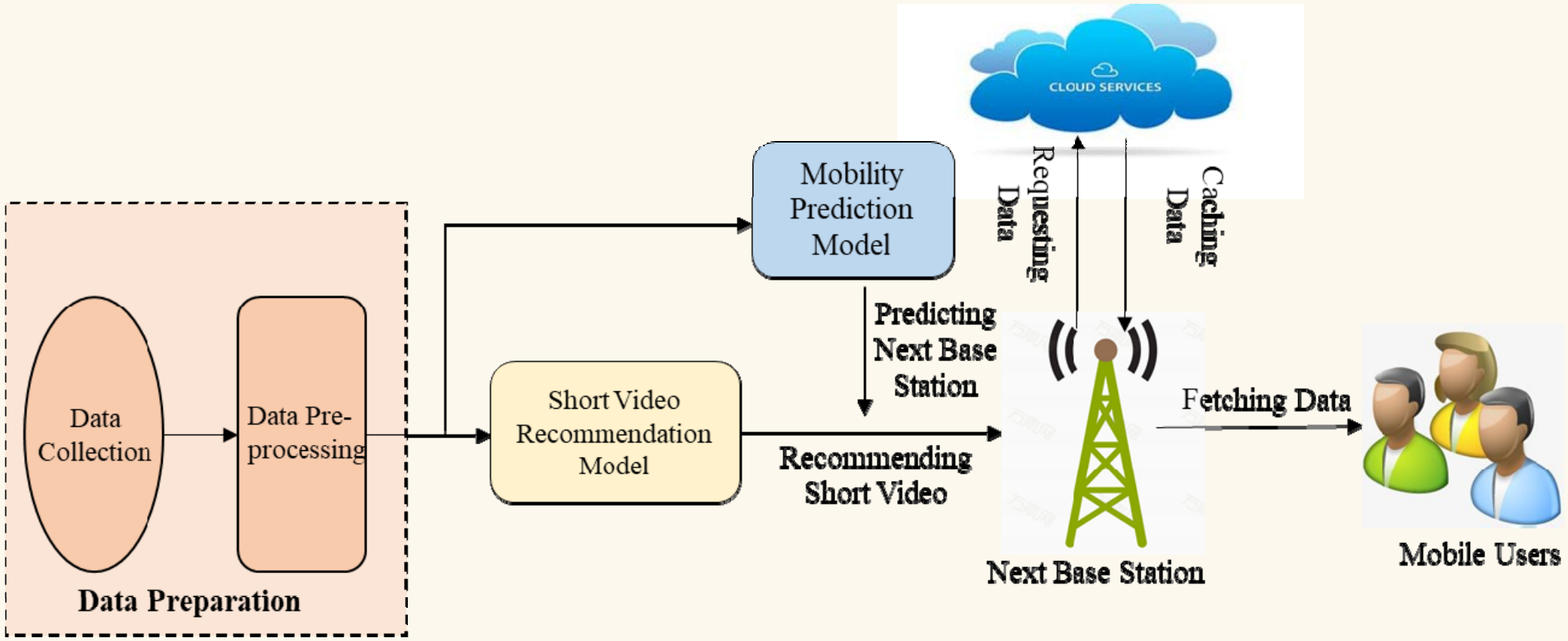
Watching the preferred video

Accurately recommend the interested video clips to the user

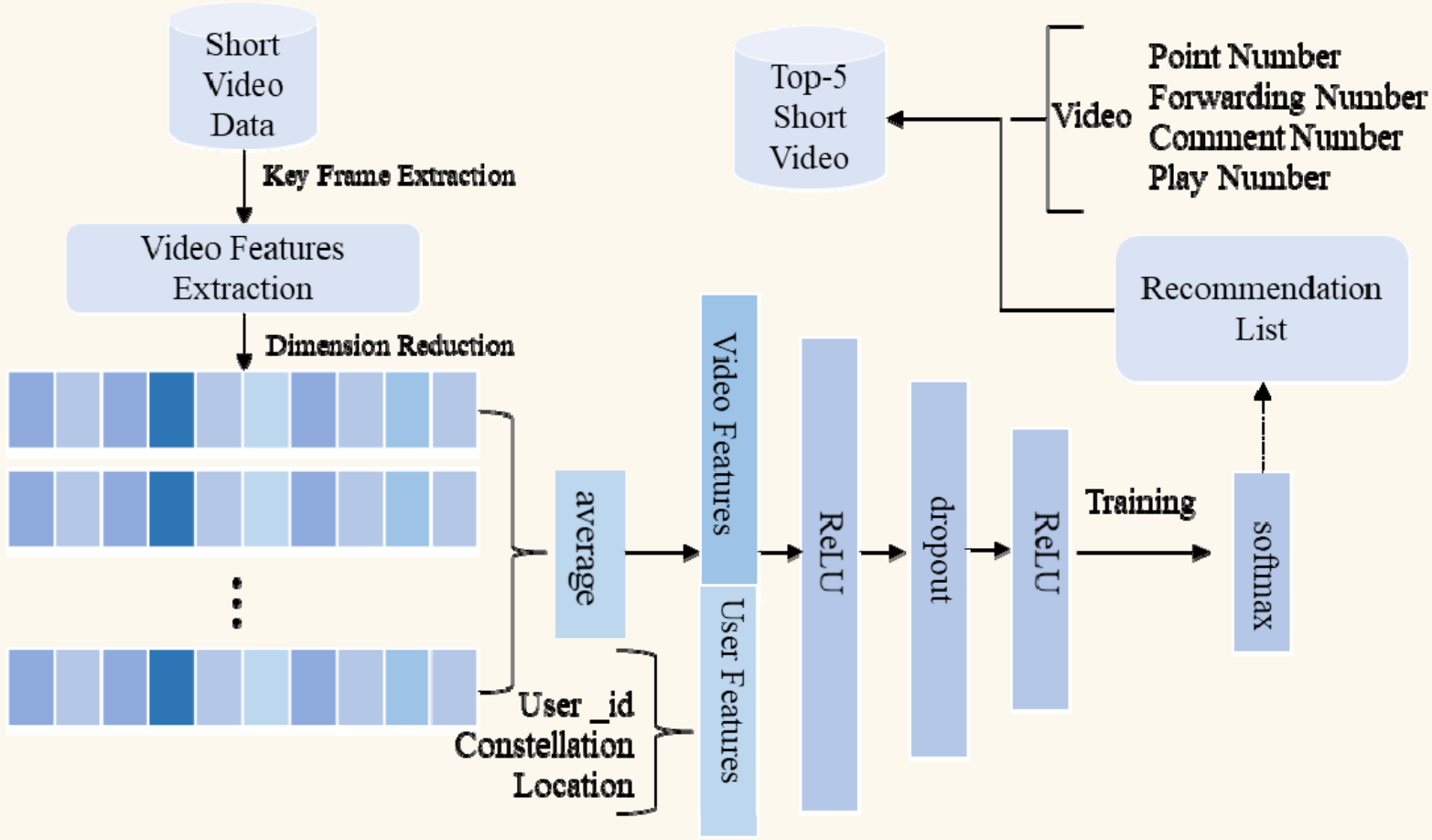
Watching it in time

Accurately prefetch the recommended video clips to the BS where the user would connect to

Overall system architecture



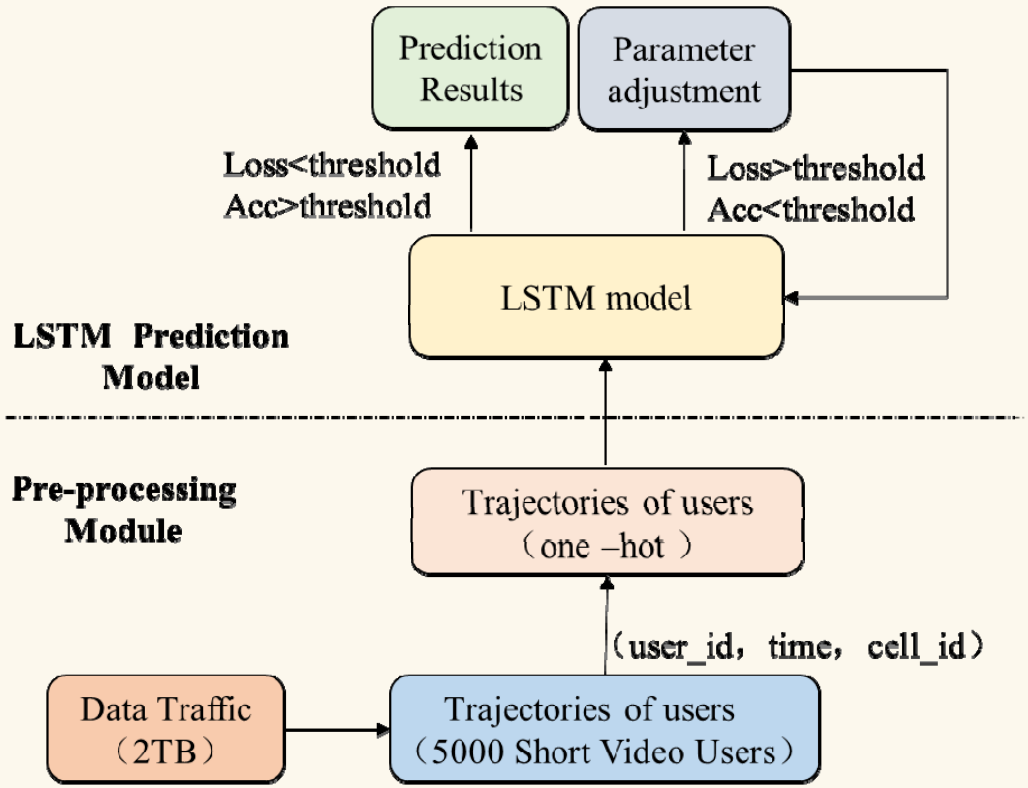
Short video recommendation



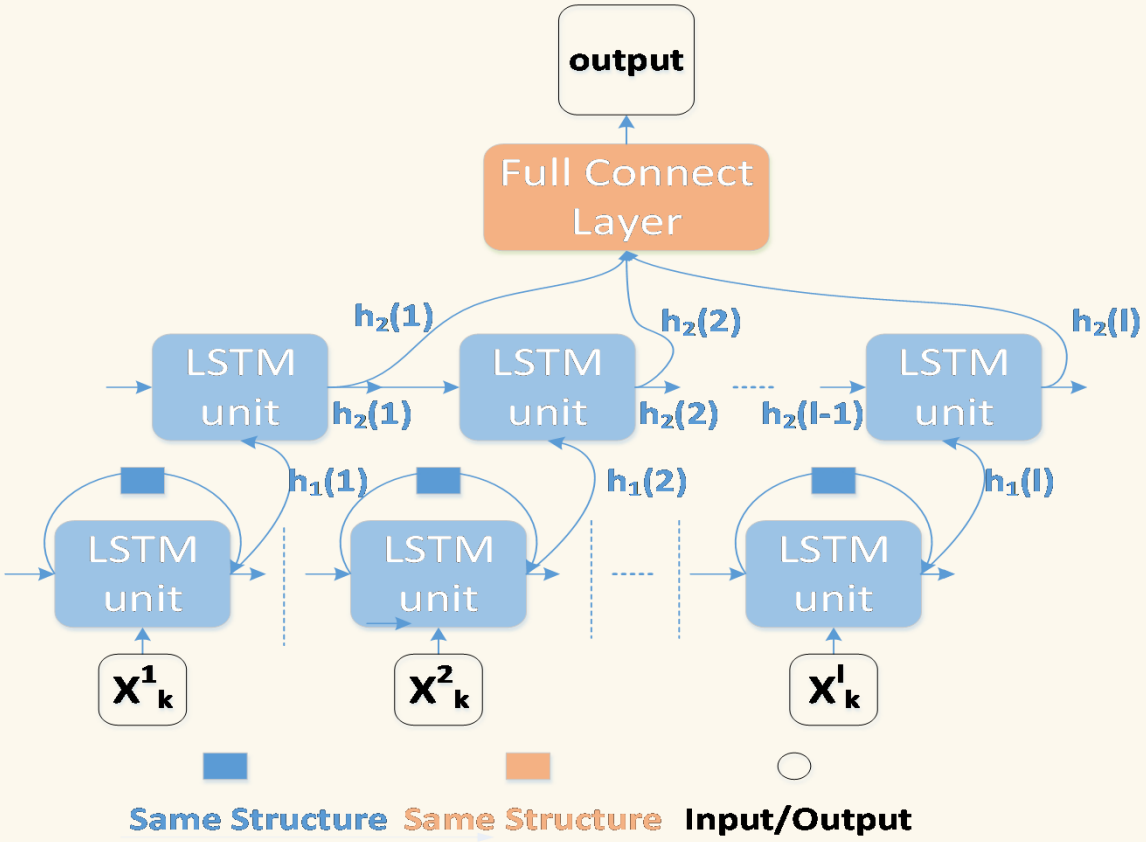
Google Inception: YouTube-8M: A Large-Scale Video Classification Benchmark. CVPR'16.

Mobility prediction

Mobility prediction structure



LSTM model



Evaluation: Dataset and Preprocessing

User mobility trace:

5,000 users event-driven trace from one of the largest ISPs in China for one week

Contains start time, user ID, downlink speed, cell ID etc.

Short video dataset:

Crawled from Douyin of 78170 records from 233 users

Includes user ID, avatar, nickname, location, constellation; video ID, release time, associate user ID, # of likes\comments\forwardings\shares and video file

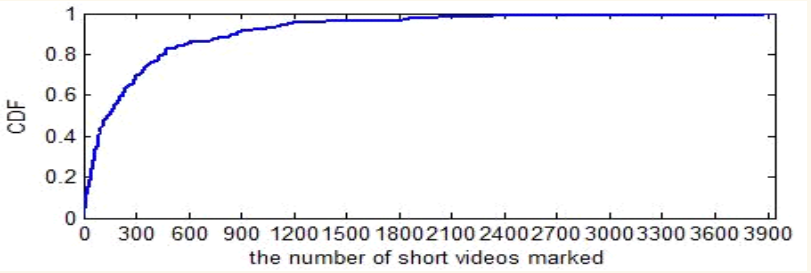
70% for training, the rest 30% for testing

Sample balancing:

Classify short videos using YOLO3 with COCO coefficients

Obtain user preference of types according to her mark

Randomly extract unmarked video from the least liked types as negative examples



Results

Performance of short video recommendation using different loss function and number of nodes

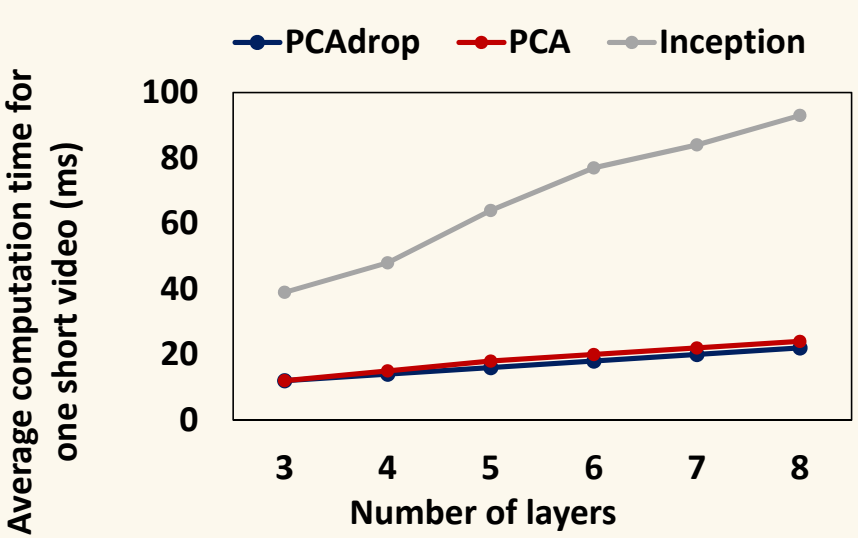
		Acc	F1
Loss function (nodes=50)	Mean absolute error	0.664	0.798
	Cross entropy	0.667	0.808
	Mean square error	0.688	0.815
Nodes (Loss function = MAE)	10	0.67	0.808
	30	0.69	0.817
	50	0.664	0.798
	128	0.68	0.814
	160	0.65	0.791
	256	0.66	0.796

Recommendation results using different neural networks layers

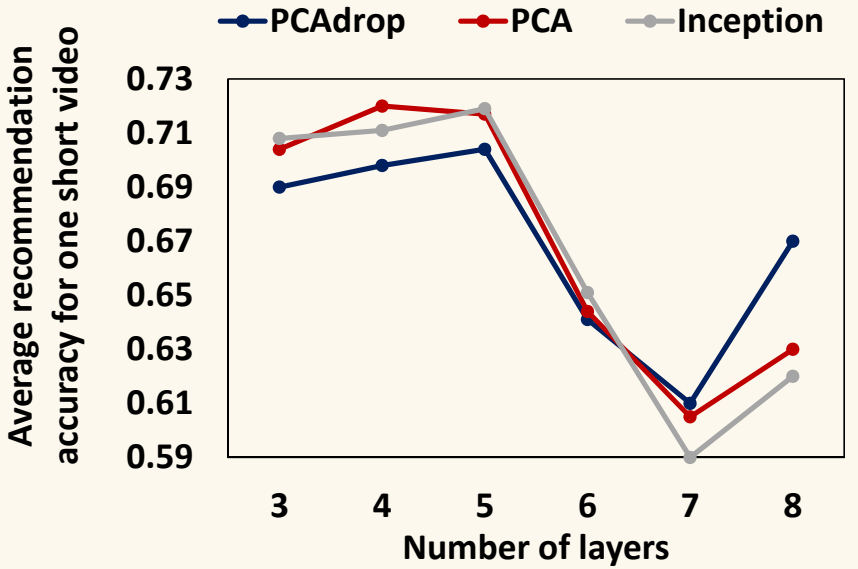
Layers	Acc	F1
3	0.69	0.817
4	0.698	0.822
5	0.704	0.826
6	0.641	0.77
7	0.61	0.75
8	0.67	0.8

Results

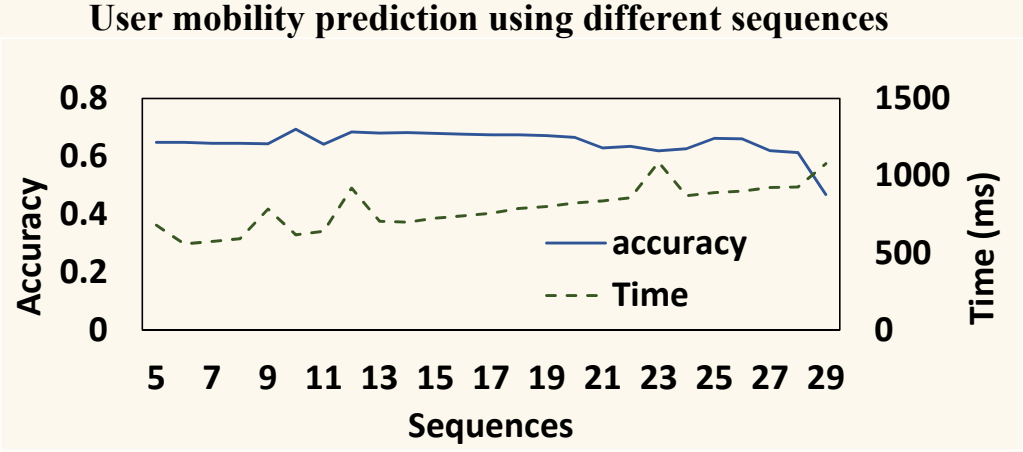
Comparison to original Inception structure w.r.t. computation time

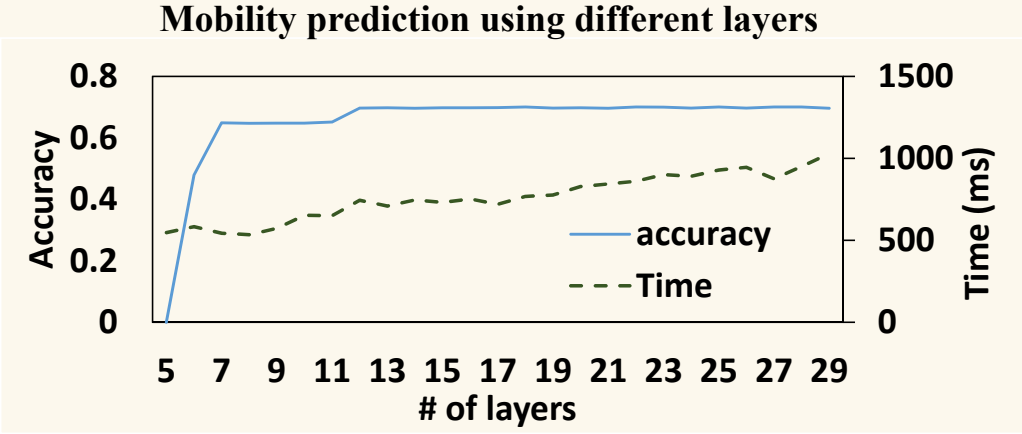


Comparison to original Inception structure w.r.t. accuracy



Results


User mobility prediction using different sequences
0.8
0.6
0.4
0.2
0
Accuracy
1500
1000
500
0
Time (ms)
accuracy
Time
5 7 9 11 13 15 17 19 21 23 25 27 29
Sequences


Mobility prediction using different layers
0.8
0.6
0.4
0.2
0
Accuracy
1500
1000
500
0
Time (ms)
accuracy
Time
5 7 9 11 13 15 17 19 21 23 25 27 29
# of layers

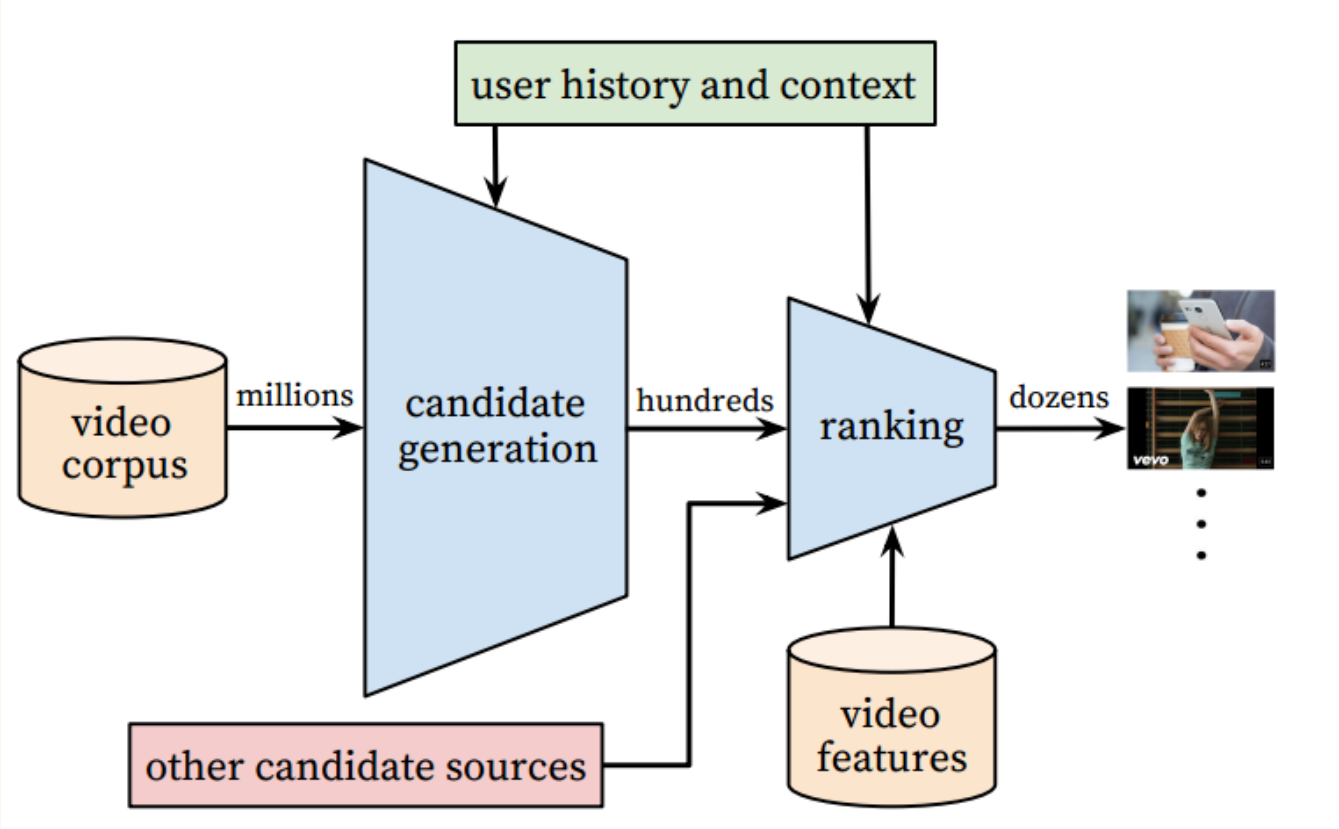
Conclusion

- A two-stage recommendation and prefetching scheme for short video application in mobile commuting scenario
- Trace-driven analysis to evaluate the accuracy and efficiency of the proposed approach

Thank you for your time!
Q&A

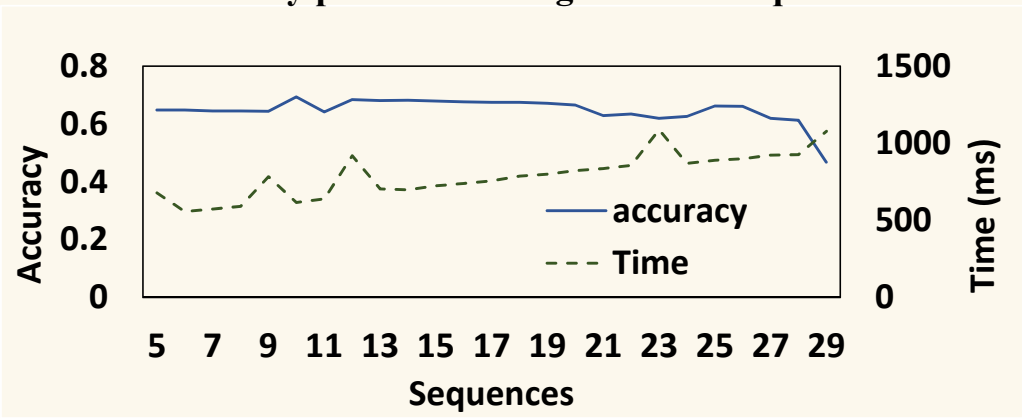
Related work

Deep Neural Networks for YouTube Recommendations，RecSys' 16

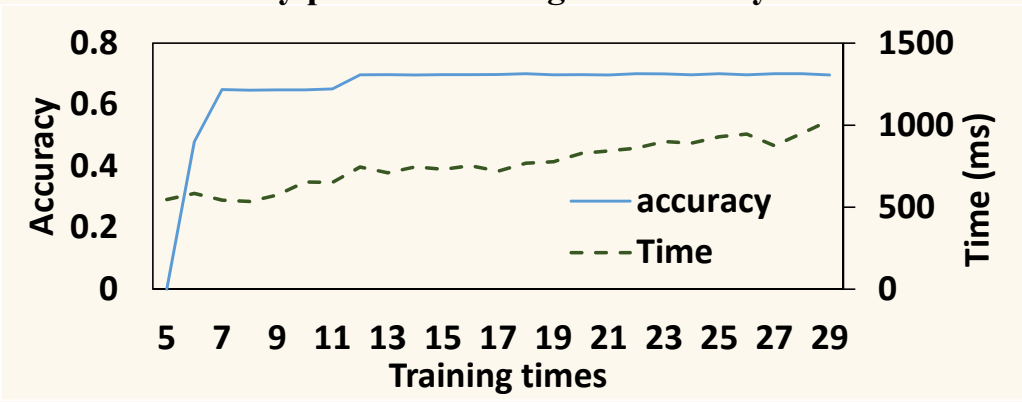


Results

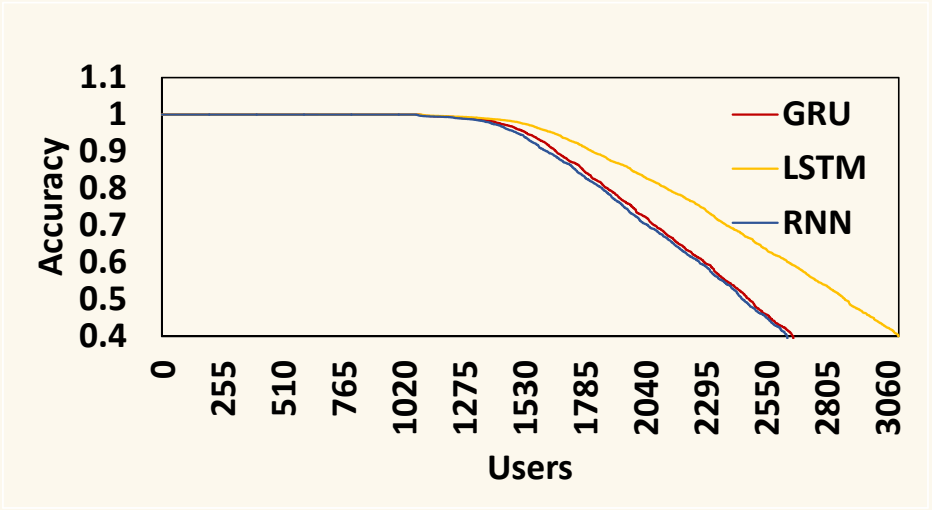
User mobility prediction using different sequences



Mobility prediction using different layers



Comparison of user mobility prediction algorithms



Feasibility Assessment

We test the network condition in commuting case for both subway and bus. In our experiment, the network condition is very unstable.

the highest download rate can reach 7.5Mbps
the lowest is only 4.8Mbps, and the average download rate is 5.15 Mbps
The maximum size of the crawled short video files is 5.2MB, and is on average 2.5MB
While the crawled video length is between 7.2s to 25.3s
therefore, in the worst case, the overall recommendation and transmission time of the short videos is 12ms+5.2*8/4.8s=8.7s

It means that for the worst case, the user would wait for 8.7s-7.2s=1.5s to get her preferred short video clips using our prefetching scheme, which is quite tolerant compared to the 8.7s waiting time without prefetching.