Predicting Online Video Advertising Effects with Multimodal Deep Learning

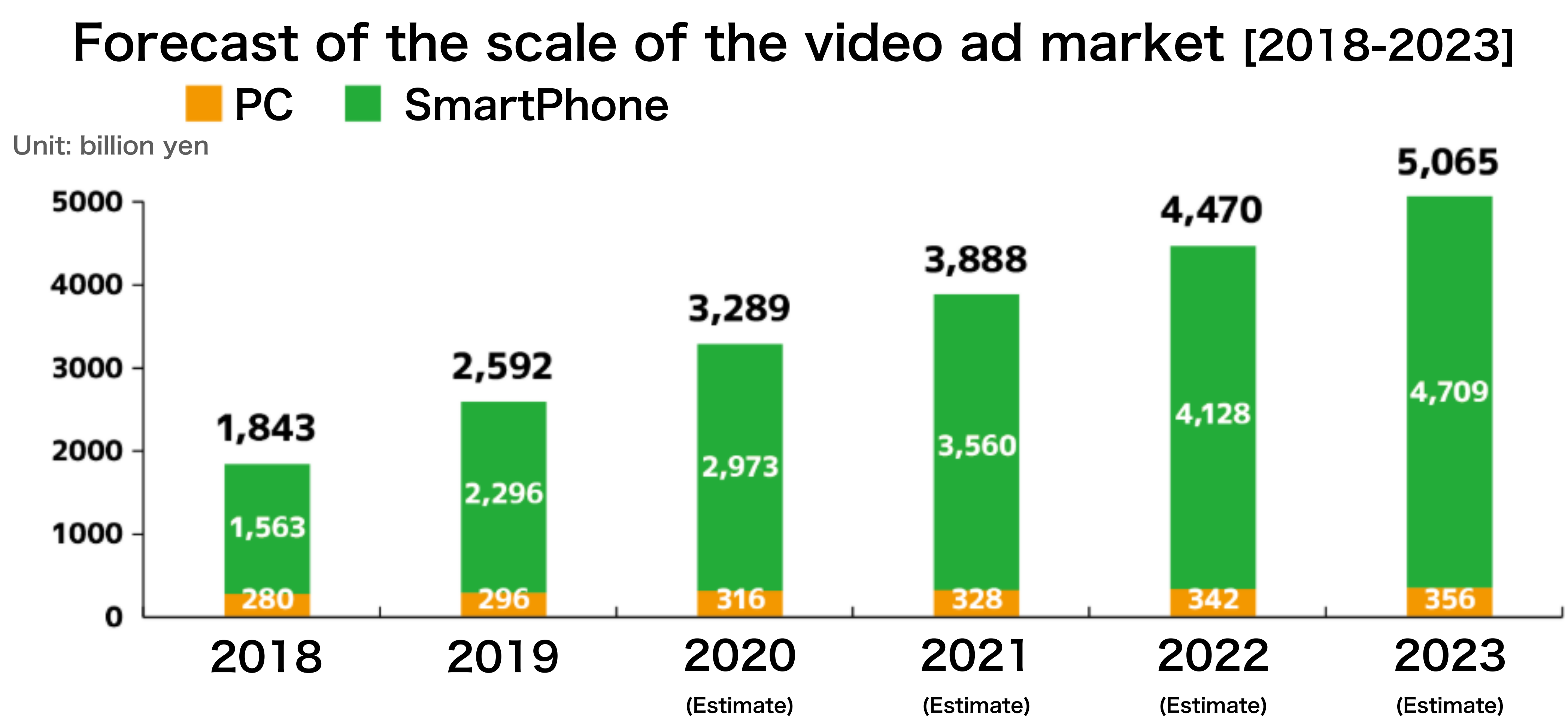
Jun Ikeda[*], Hiroyuki Seshime[†],
Xueting Wang[*] and Toshihiko Yamasaki[*]

[*] The University of Tokyo, Tokyo, Japan.
[†] Septeni Co., Ltd. Tokyo, Japan.

ICPR2020

Motivation

- Effective video ads are getting more important as the market is expanding.



Source: CyberAgent, Inc. (https://www.cyberagent.co.jp/news/detail/id=24125)

Purpose of this work

- Predict Crick Through Rate (CTR) of online video ads.
  - Assist ad designers creating more effective ads.
  - Enable designers to select the most effective ads beforehand.

Online video ad

$$\text{CTR} \left(= \frac{\text{Number of clicks}}{\text{Number of impressions}} \right)$$

Data

- Online video ad data
  - Actually used in a business by Septeni Co., Ltd.
  - Distributed on Facebook and Instagram since January 2018 until December 2019.
  - Consist of ad videos, 16 kinds of metadata, and 5 kinds of text data.

The partial example of metadata and text data.

Key	Value	Key	Value
Month	10	Target age min	13
Genre	Game	Target age max	65
Sub genre	RPG	Target cost	4500000
web/app	App	Title	The world of “Seven Deadly Sins” ….

Dataset

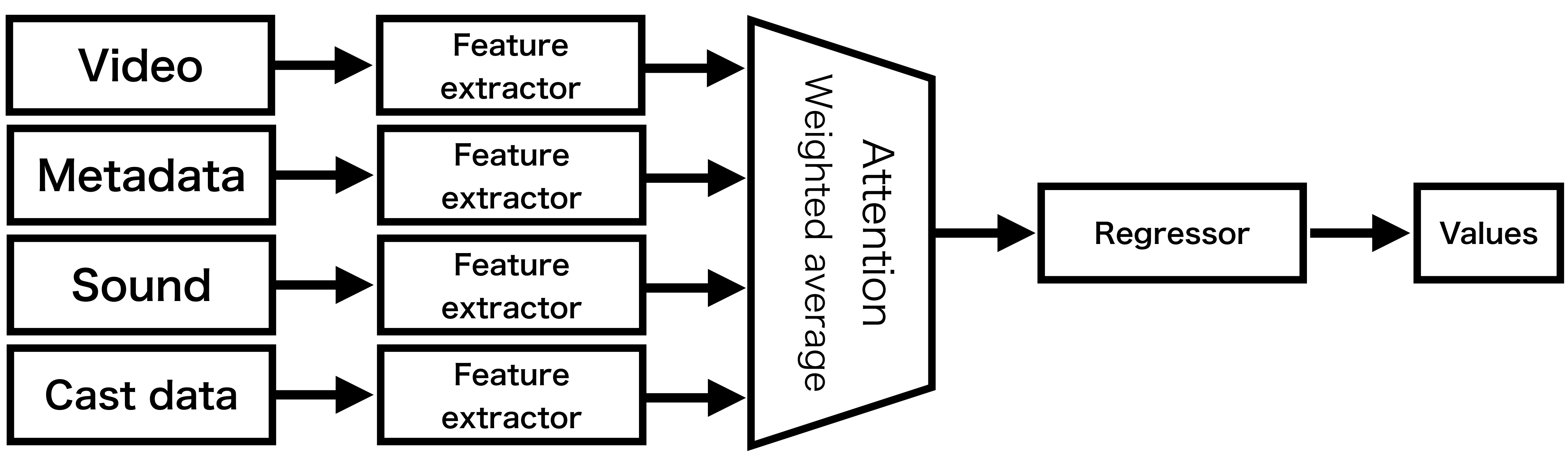
- Data split into train, val and test dataset.
  - The ads in the validation and test datasets should be newer than those in the training dataset.
- Several ads share the same video.
  - Ads with the same video content shouldn't be separated between datasets.

The numbers of data

	Train	Val	Test
	Older ads ←	→	Newer ads
Ads	80,771	8,061	9,655
Unique videos	20,618	2,032	2,945

Related study

- Prediction of TV commercial impressions. [Nakamura et al.]
  - Predict four impressional and emotional effects of 15s TV commercials, using video, metadata, sound and cast data.



Problem of Nakamura's model

- Differences between research targets, TV commercials and online video ads.

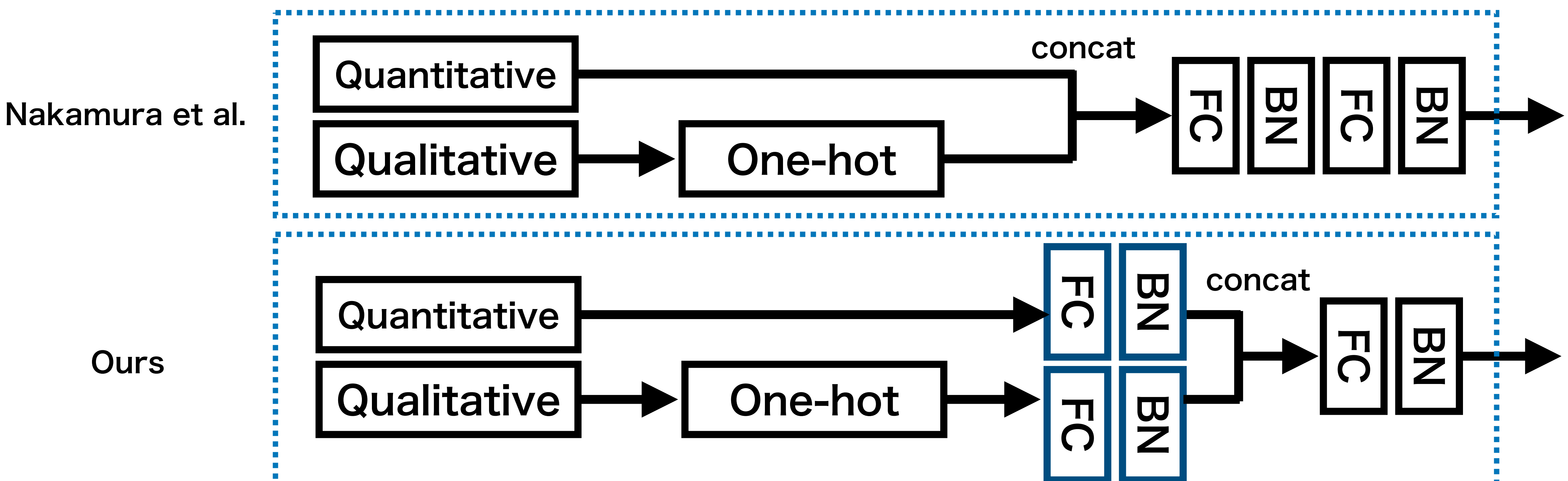
Differences

	Nakamura et al	Ours
Data	TV commercials	Online video ads
Kinds of data	Video, metadata, sound, cast data	Video, metadata, text
Data features	—	Many similar data Several numerical metadata

Proposed method

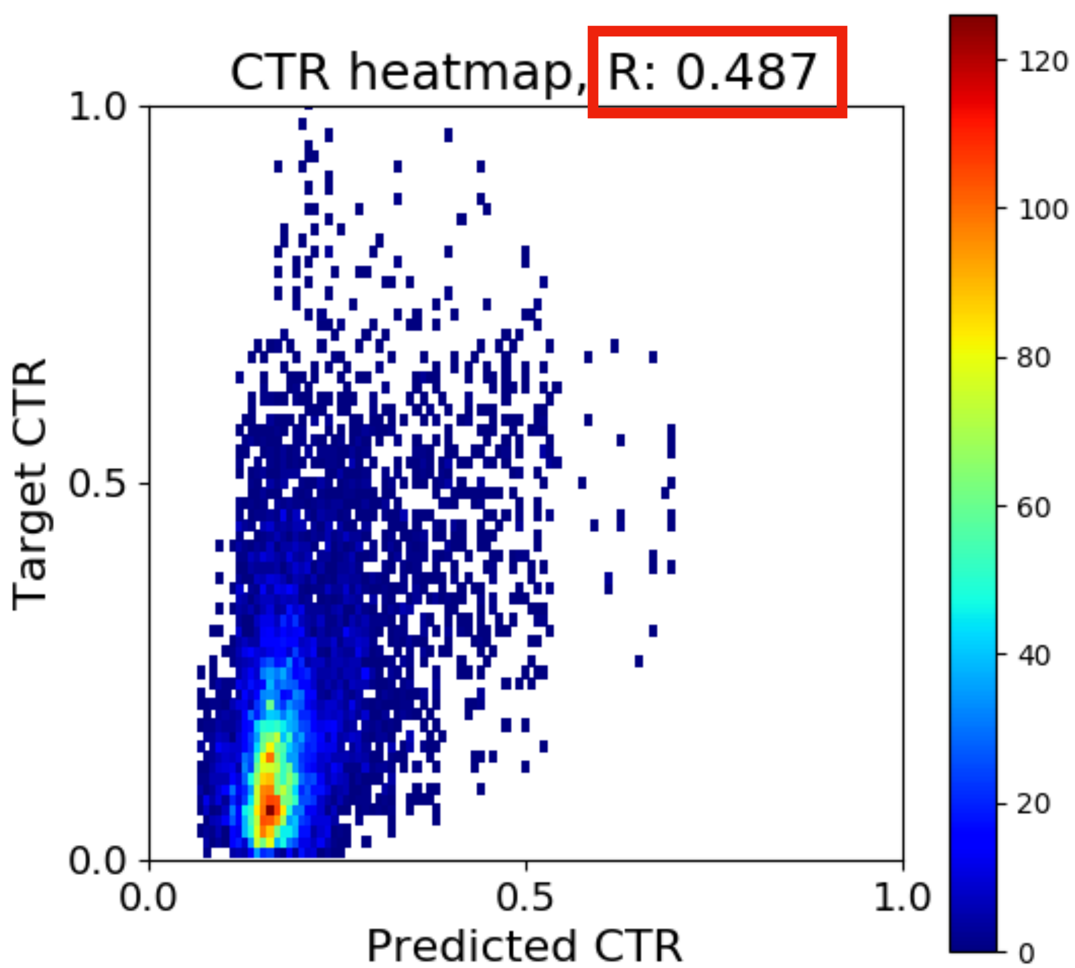
- Improve metadata feature extractor.
- Suppress overfitting using batch normalization and dropout.
- Input embedded text data.

Metadata feature extractor
(FC: fully connected layer, BN: batch normalization)

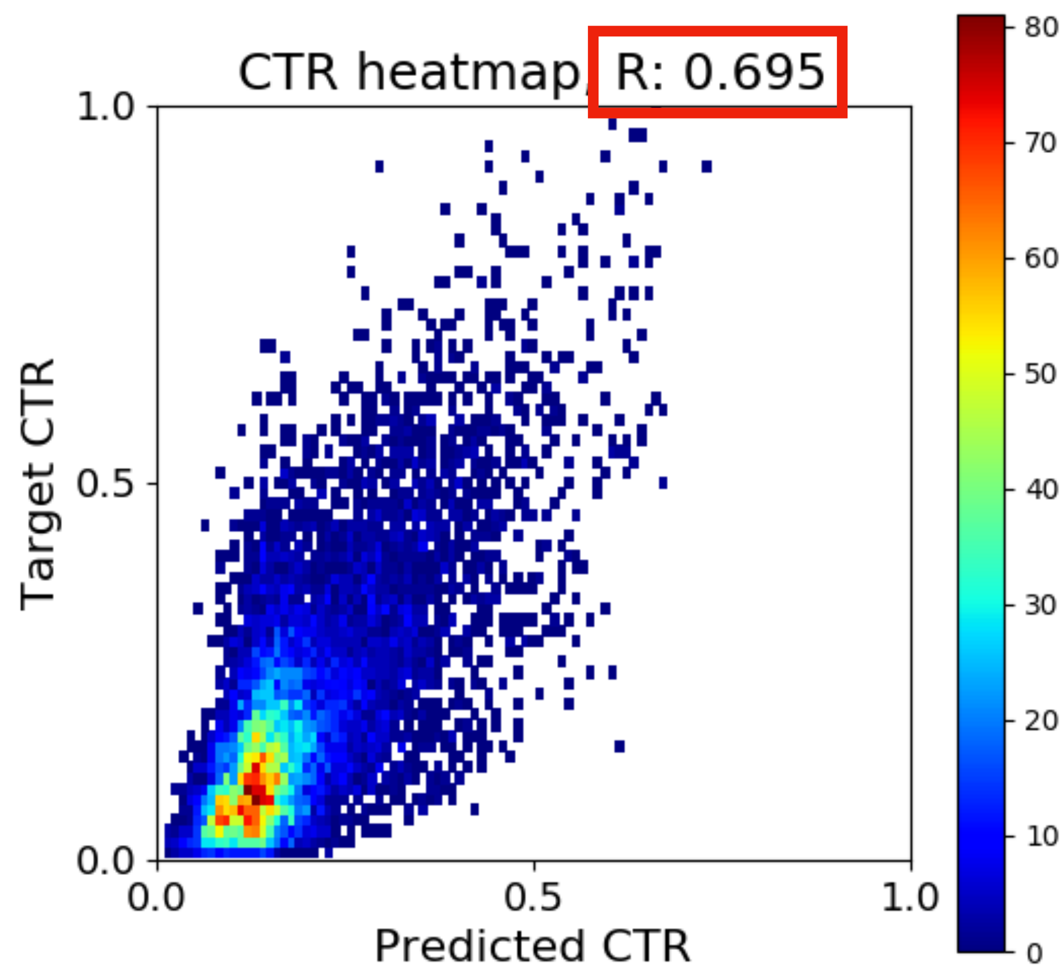


Results

- Achieve a higher correlation coefficient (0.695) than Nakamura et al. (0.487).



Nakamura et al.



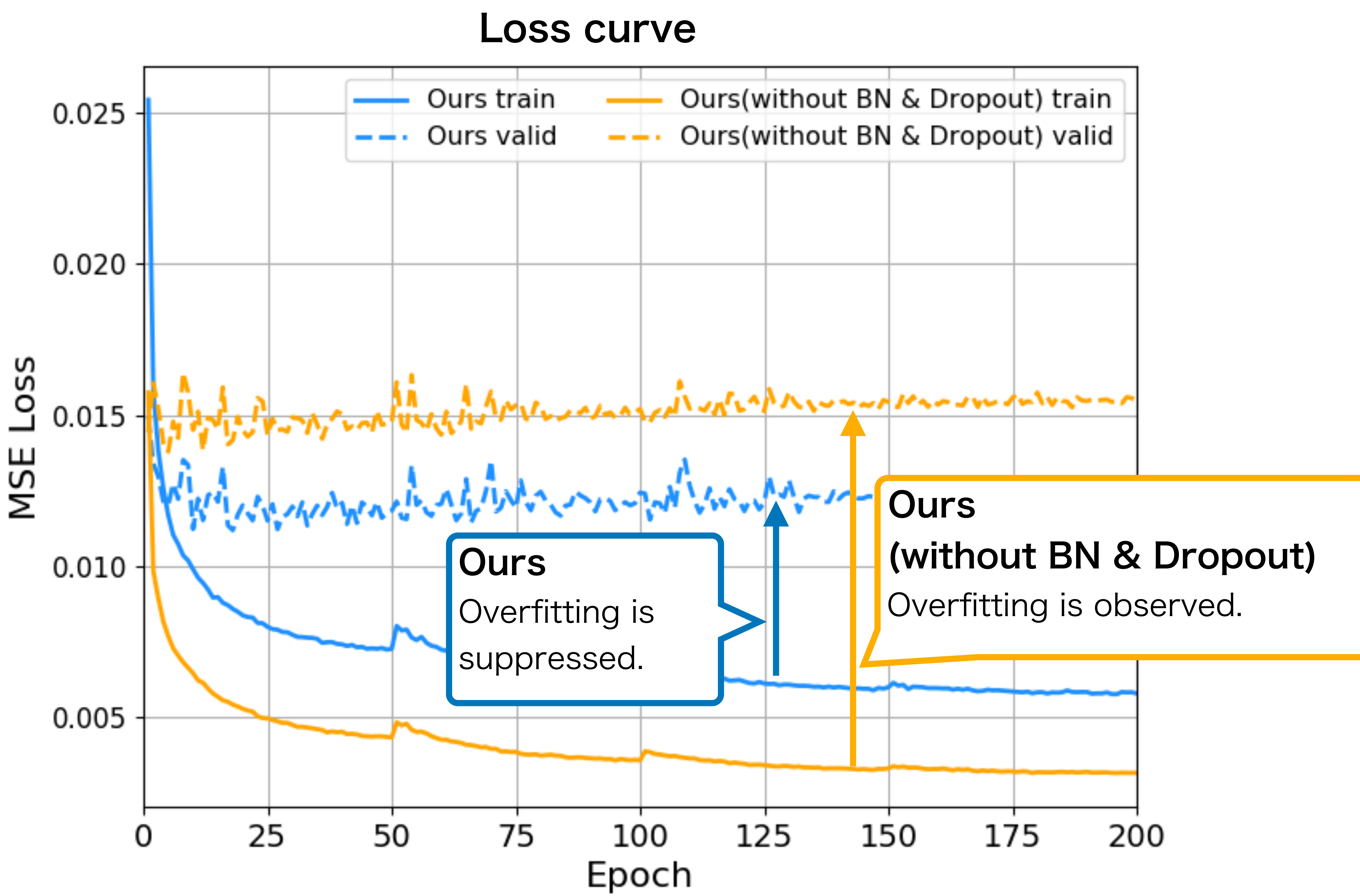
Ours

Results

- Ablation studies demonstrate the contribution of our proposals.

Method	Input			Meta feature extractor	Suppress overfitting	Metrics	
	Video	Meta	Text	Improved	BN & Dropout	RMSE↓	R↑
[Nakamura et al.]	✓	✓				0.130	0.487
Ours(without improved extractor)	✓	✓	✓		✓	0.126	0.540
Ours(without text input)	✓	✓		✓	✓	0.109	0.684
Ours(without BN & Dropout)	✓	✓	✓	✓		0.121	0.598
Ours	✓	✓	✓	✓	✓	**0.107**	**0.695**

Results



Loss curve
Ours train
Ours valid
Ours(without BN & Dropout) train
Ours(without BN & Dropout) valid
0.025
0.020
0.015
0.010
0.005
MSE Loss
0
25
50
75
100
125
150
175
200
Epoch
Ours
Overfitting is suppressed.
Ours
(without BN & Dropout)
Overfitting is observed.

Conclusion

Purpose

Predict CTR of online video ads.

Related Study

Predict effects of TV commercials, which have different features of data from online video ads.

Proposed method

Improve metadata feature extractor.

Suppress overfitting.

Input text data embedded by Doc2Vec.

Results

Achieve a correlation coefficient as high as 0.695.