CTR PREDICTION SYSTEM BASED ON WIDE & DEEP LEARNING (COMBINED WITH GBDT)

Group 10
A10715003 Xuanjun Chen
M10415014 Zheng Wei

OUTLINE

- ➤ Introduction
- ➤ Architecture
- ➤ Experiment
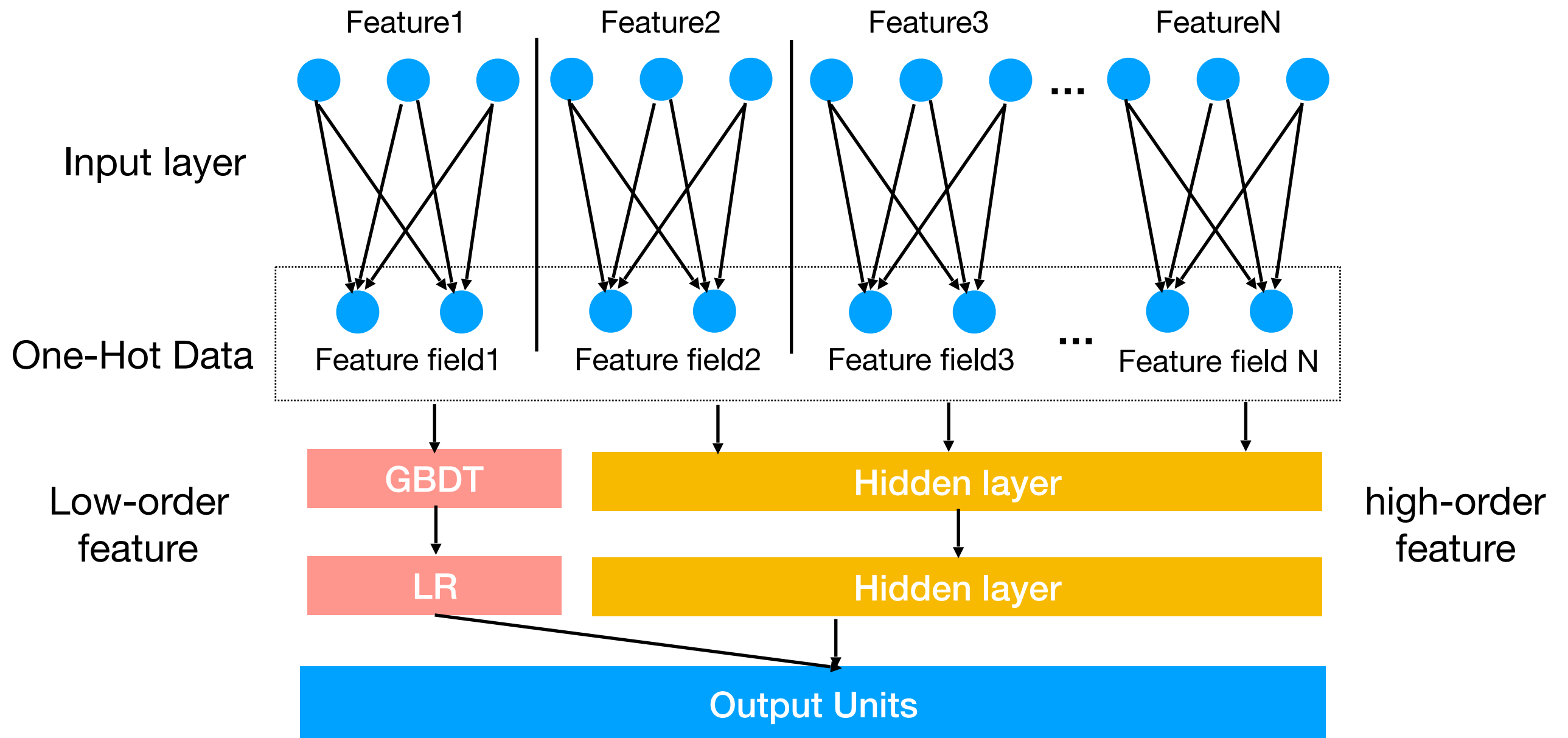- ➤ Future Work
- ➤ Conclusion

INTRODUCTION

Problem

➤ Click-through Rate prediction

➤ Based on Wide & Deep learning model

➤ Wide component can't deal with very high dimensions.

Our solution

➤ LDG model combines Wide & Deep model with Gradient Boosting Decision Tree (GBDT)

ARCHITECTURE

EXPERIMENT

➤ Dataset: Avazu Dataset

➤ Max depth of decision tree = [1,2,3,4,5,6,7,8,9]

➤ Data scales = [10k, 20k, 30k, 40k, 50k, 60k, 70k, 80k]

➤ Training time

➤ Other :

   ➤ Train / Validation = 0.95

   ➤ Early stopping

   ➤ Epoch = 100

   ➤ Batch size = 20

EXPERIMENT

Layer (type)	Output Shape	Param #	Connected to
input_2 (InputLayer)	(None, 929)	0	
dense_2 (Dense)	(None, 20)	18600	input_2[0][0]
batch_normalization_1 (BatchNor	(None, 20)	80	dense_2[0][0]
activation_1 (Activation)	(None, 20)	0	batch_normalization_1[0][0]
dense_3 (Dense)	(None, 20)	420	activation_1[0][0]
batch_normalization_2 (BatchNor	(None, 20)	80	dense_3[0][0]
activation_2 (Activation)	(None, 20)	0	batch_normalization_2[0][0]
dense_4 (Dense)	(None, 20)	420	activation_2[0][0]
batch_normalization_3 (BatchNor	(None, 20)	80	dense_4[0][0]
input_1 (InputLayer)	(None, 519)	0	
activation_3 (Activation)	(None, 20)	0	batch_normalization_3[0][0]
dense_1 (Dense)	(None, 1)	520	input_1[0][0]
dense_5 (Dense)	(None, 1)	21	activation_3[0][0]
add_1 (Add)	(None, 1)	0	dense_1[0][0] dense_5[0][0]
activation_4 (Activation)	(None, 1)	0	add_1[0][0]

Total params: 20,221
Trainable params: 20,101
Non-trainable params: 120
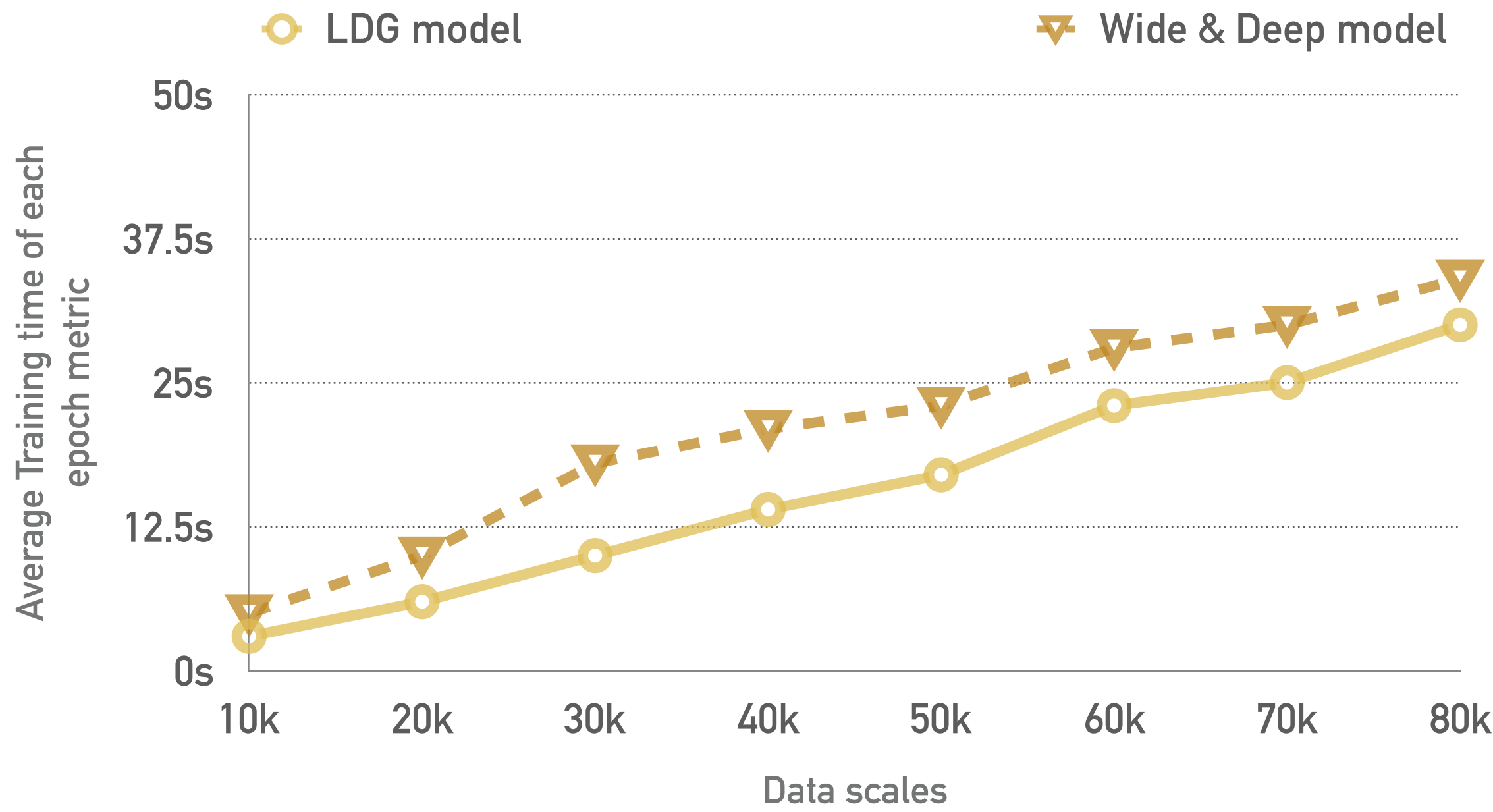
Fig. 2. Hybrid model detail

EXPERIMENT

Comparison of decision trees with different maximum depths

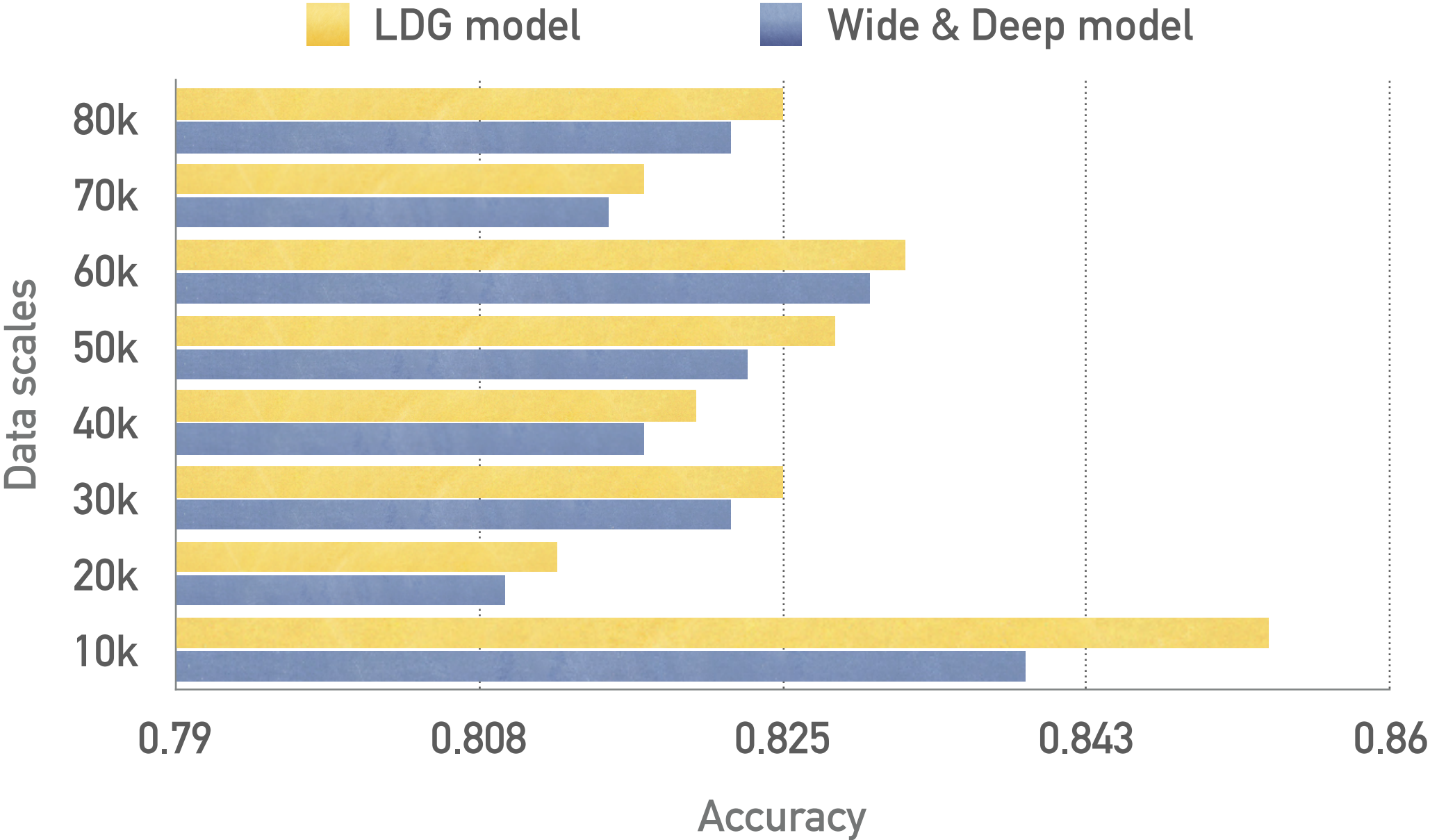○ LDG model   ▽ Wide & Deep model
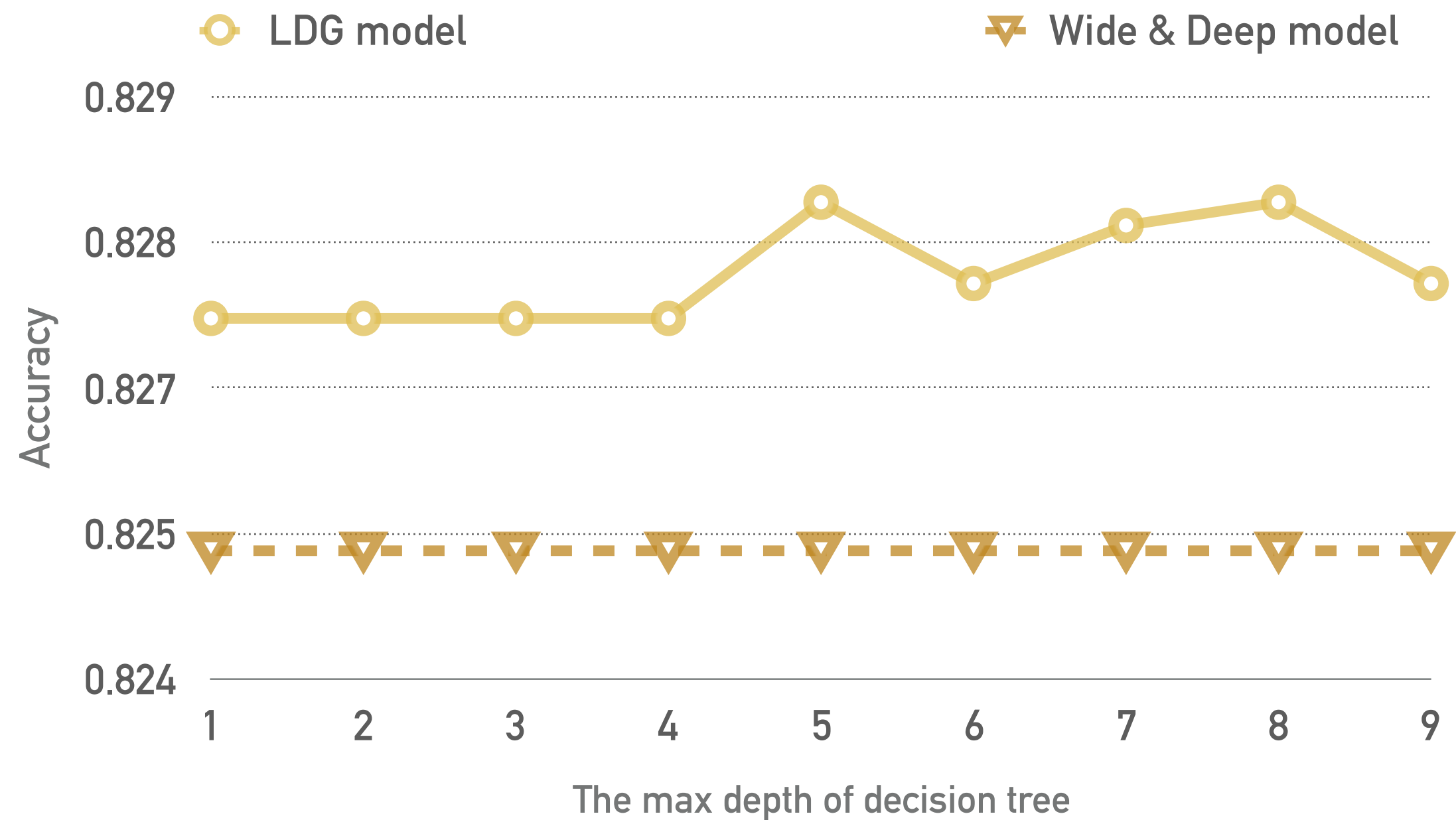
EXPERIMENT

Comparison of training time

○ LDG model   ▽ Wide & Deep model

FUTURE WORK

➤ Time series features are important for CTR prediction system

➤ Algorithm limit (GBDT)

➤ Hardwares limit

CONCLUSION

➤ We find that Gradient Boosting decision tree definitely improves the capabilities of Wide & Deep model by reduce the input dimensions of wide linear model.

➤ LDG model spend less time than Wide Deep learning model.

➤ LDG model only performs better than Wide & Deep learning model in small-scale dataset.

➤ The maximum depth of the decision tree will only affect the accuracy slightly.