

A decorative network diagram in the top-left corner of the slide. It features a complex web of interconnected nodes and lines. The nodes are represented by circles of varying sizes and colors, including light gray, dark gray, and blue. Some nodes are highlighted with a blue outline. The lines connecting the nodes are thin and gray, creating a mesh-like structure.

RAKUTEN CASE STUDY

Rakuten Advertising Case Interview

Monday 25th March, 2024

Fanxiang ZENG

A decorative network diagram in the bottom-right corner of the slide. It features a complex web of interconnected nodes and lines. The nodes are represented by circles of varying sizes and colors, including light gray, dark gray, and blue. Some nodes are highlighted with a blue outline. The lines connecting the nodes are thin and gray, creating a mesh-like structure.

Data structure

Data structure



Numerical columns

- **imps** : number of impressions
- **visibility** : visibility of the ad
- **width** : width of the ad
- **height** : height of the ad
- **clikcs** : number of clicks

Categorical features

- **date**
- **country_ref**, **domain**, **device**
- **IDs** : **zone_id**, **media_id**, **advertiser_id**, **campaign_id**, **ad_id**

Data structure



Dimension of dataset

- train data : (5395051, 14)
- test data : (231201, 13)

Target variable

- clicks : number of clicks
- ctr : click-through rate

Data pre-processing

Data pre-processing



Handle missing values

- **Numerical Columns** : fill missing values with the median of the column
- **Categorical Columns** : fill missing values with 'Unknown'

Date

- extract day, month, year, weekday
- drop the original date column

Feature scaling and encoding

- Label encoding for categorical columns
- Standardize numerical columns

Model

How to predict click-through rate?

- Predict the click number using regression model and then calculate the CTR by dividing the predicted click number by the number of impressions
- **Tree-based model** : capture non-linear relationship
- **Deep learning model** : learn complex patterns in the data

Model selection



Model selection

- MLP : Multi-layer Perceptron ✗
- DCN : Deep & Cross Network ✓

Evaluation metric

$$-\frac{\sum_i \text{clicks}_i \times \ln(\hat{y}_i) + (\text{imps}_i - \text{clicks}_i) \times \ln(1 - \hat{y}_i)}{\sum_i \text{imps}_i} \quad (1)$$

- **clicks** : number of clicks
- **imps** : number of impressions
- \hat{y} : predicted click-through rate
- **BCE loss** : binary cross-entropy loss

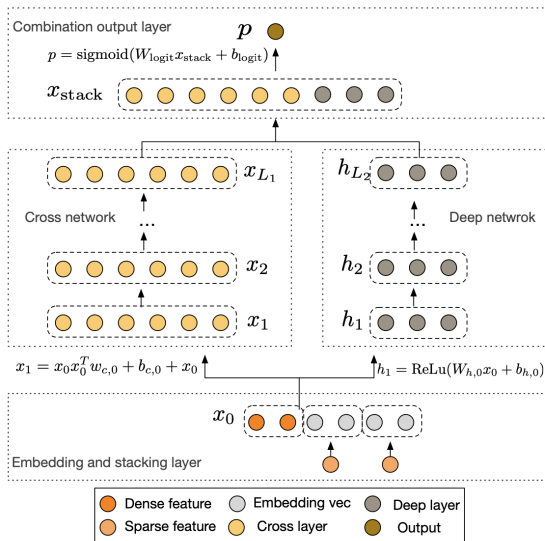
Model Description



Deep & Cross Network

- Learn tabular data by transforming categorical features into dense vectors through embeddings
- **Cross Network** : capture the relationship between features with a series of cross layers
- **Deep Network** : fully connected layers

Model Description



Model training



Model training

- Optimizer : Adam
- Learning rate : 0.001
- Batch size : 128
- Epochs : 5 or 3

Results

Results



cf. Notebook and .csv

Results



Ratio of non zero ctr

- BCE CTR label : 1676703 / 2019086
- BCE CTR click label : 1439133/ 2019086
- MSE CTR label : 319 / 2019086
- Clic predict CTR : 1105119/ 2019086

Evaluation metric

- BCE CTR label : 0.05838478356599808
- BCE CTR click label : 0.08586359024047852
- MSE CTR label : 0.09446008503437042
- Clic predict CTR : 0.4865149259567261

Conclusion

Thank you for your attention!