COMP0124 Multi-Agent Artificial Intelligence Group Project

Lecturer: Prof. Jun Wang
jun.wang@cs.ucl.ac.uk

TAs: Ying Wen
{ying.wen.15}@ucl.ac.uk

Submission Deadline: 23:55 08/03/2019

This group project consists of one group report, code and testing set result file,

worth 50 marks (50% of your overall score)

1. SUMMARY

In this assignment, you are required to work on an online advertising problem. You will help advertisers to form a bidding strategy in order to place their ads online in a real-time bidding system. You are required to train a bidding strategy based on a provided advertising impression training set. This project aims to help you understand some basic concepts and write a computer program in real-time bidding based display advertising. The performance of the model trained by your team, which is either a combination of the individually developed models or the best performing individually-developed model, will be (mainly) evaluated on the Click Number achieved on a provided testing set. For you to properly evaluate the performance of each of your models before that, a benchmark click-through rate on the validation set will be provided.

Before the final submission, you are also given the opportunity to hand in the preliminary result of your team's model on the testing set via the Leaderboard Website or API (API Document), which allow you to compare the performance to that of your peers. And the preliminary results may usually limit 5 submissions daily. Also we have a frequent Q&A list can be found here link. Note that the online prelimiary submision is new this year and you might encourter some technical problem. Should you have any question, feel free to contact group project TA: Ying Wen (email).

2. RULES

You are expected to work in a team of 3 people, and you can sign in the team through this google sheet, by next Monday (28/01/2019). The team will jointly write a group report.

2.1 Submissions (In Moodle)

- Final submission: 23:55 08/03/2019. Please upload your submission through Moodle as a zip file, and rename the file as that: Group_xx.zip (e.g., Group_10.zip). For the final submission, the file should include the report in PDF format (including the names of all members of your team, the report should be named as group_xx_report.pdf), and your team's predictions for the provided test set. Please upload your code to GitHub and provide a link to it in the reports (again for both reports). Note that we only take the results from the final submission to calculate the winning group (top three group would get additionl marks, provided that the overall group marks are no more than 50).

2.2 The Report

As a group, you are required to submit one *Group Report* (one per group), which in total worths 50 marks. The rest 50 marks will come from the individual coursework, which should be released later.

Given in Section 3.2, the reports shall summarise what your group did and how you addressed the posed questions as described in the problem description.

You will use the ACM Proceedings style LATEX files for writing your reports. You can find out more about these styles as well as LATEX templates on the ACM website. Note that to make your report looks better, you could simply add the following lines before `\begin{document}` to remove the copyright notice:

```
\makeatletter
\def\@copyrightspace{\relax}
\makeatother
```

For the *Group Report*, your submission should be 5-7 pages, with maximum 1 additional page for references. Images and tables also count against the maximum. Rules written here trump all the non-formatting instructions in the ACM style files. Your report should address all the problems in Section 3.2 and follow the format from Section 3.4.

2.3 Q&A and Office Hours of TAs

Email us if you have any question. Alternatively, if you would like to get answered face to face, please come to the office hours of this module as below:

- Time: 12:00-13:00, 05/19/26 Feb 19, 05/12/19 Mar 19.
- Location: Hub Level 3, 66-72 Gower Street. (Except 12 Mar 19, we will move to Hub Level 1.)

Note, please send an email to us in advance before you attend an office hour, making sure TAs are available.

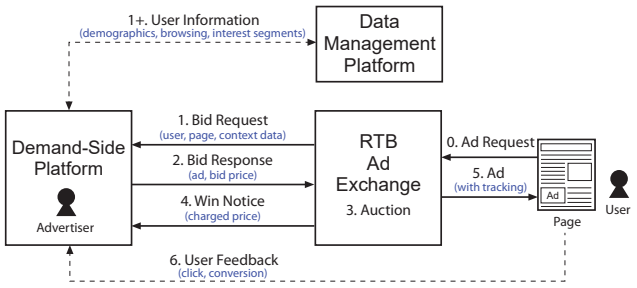

Figure 1: A brief illustration of the interactions between user, ad exchange and the advertiser's DSP bidding agent.

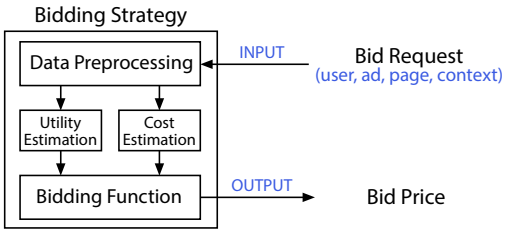

Figure 2: A bidding strategy can be abstracted as a function mapping from the given bid request (in a high dimensional feature space) to a bid price (a non-negative real or integer number).

3. THE PROJECT

In display and mobile advertising, the most significant evolution in recent years is the employment of the so-called Real-Time Bidding (RTB) mechanism to buy and sell ads. RTB essentially facilitates buying an individual ad impression in real time, automatically triggered by a user's visit. Although other types of auctions, such as the first price auction, are also popular, RTB exchanges typically employ the second price auction model. In this type of auction, the bidder winning the auction pays the price of the second highest offer. A summary of the process is provided in Figure 1. In this assignment you will be training your own bidding strategy models on a given impression dataset, as described in detail next.

3.1 Data Description

The dataset includes a training set, a validation set, and a testing set. The dataset for the task can be downloaded from this Onedrive Sharing Link.

This data comes in CSV format, the first line in the file containing the header formatted as described in Table 1. As the testing set is used for final evaluation purposes, it does not contain the three fields: 'bidprice', 'payprice' and 'click'. Note that that all number related to money (e.g., bid price, paying price and floor price) is Cost Per Mille (CPM), corresponding to the commonly adopted CPM pricing model. It uses the currency of CNY and the unit is Chinese fen. For instance the payprice in Table 1 means the cost is 322 Chinese fen per 1000 impressions. When someone won this impression, they had to pay 0.322 Chinese fen just for it.

You can develop your model on the training data and use the validation set to compare the different models and correct for overfitting.

3.2 Problem Description

In this assignment, you are required to train a bidding strategy based on the impressions in the training set. This objective has been divided into the following five sub-problems, and you are required to solve and report these problems in either the group or individual report, as indicated below. The concept of a bidding strategy has been visualized in Figure 2 and can also be found in the lecture notes.

3.2.1 *Problem 1: Literature Review and Data Exploration*

- 10/50 marks

(5 marks for problem introduction and liteature review sections and 5 marks for data exploration). You should do some exploratory data analysis and show your understanding of this dataset. You can start from basic statistical information, such as num Imps, num Clicks, Cost, CTR, avg CPM, eCPC. You can also include further analysis on user feedback, bidding etc. A good example can be found in the paper [1].

3.2.2 *Problem 2: Basic Bidding Strategies*

- 10/50 marks

Evaluate the following strategies on the validation set and discuss your results. Use the performance metrics as defined in Section 3.3.

- Constant bidding (Const). Bid a constant value for all the bid requests. The parameter is the specific constant bid price. Explain how you can experimentally find an optimal constant value from the training set. Provide an analysis of your finding.
- Random bidding (Rand). Randomly choose a bid value in a given range. The parameter is the upper bound and lower bound of the random bidding range. Give your experiment results of finding the optimal bounds on the basis of Winning criterion #1.
- Considering competition among homogeneous random bidding agents. Using Rand bidding above to create n (ranging from 50 to 100) number of bidding agents. give your experiment results of finding the optimal bounds on the basis of Winning criterion #2. Investigate the difference compared to the single agent case and also the impact of n.

3.2.3 *Problem 3: Linear Bidding Strategy*

- 5/50 marks

Apply CTR estimation to create a linear bidding strategy. The bid value is linearly proportional to the pCTR (predicted CTR). The formula can be generally written as $bid = base\_bid \times pCTR / avgCTR$, where the tuning parameter $base\_bid$ is the bid price for the average CTR cases. Optimise the $base\_bid$ and the CTR estimation and evaluate the performance of your model using the metrics as defined in Section 3.3. Discuss and compare your results.

3.2.4 *Problem 4: Your Non-Linear Bidding Strategy*

- 5/50 marks

Table 1: Fields in dataset

Field	Example	Supplement
click	1	1 if clicked, 0 if not.
weekday	1	
hour	12	
bidid	fdfe...b8b21	
userid	u_Vh1OPkFv3q5CFdR	
useragent	windows_ie	
IP	180.107.112.*	
region	80	
city	85	
adexchange	2	
domain	trq...Mi	
url	d48a...4efeb	
urlid	as3d...34frg	
slotid	2147...813	
slotwidth	300	
slotheight	250	
slotvisibility	SecondView	
slotformat	Fixed	
slotprice	0	
creative	hd2v...jhs72	
bidprice	399	
payprice	322	Paid price after win the bidding.
keypage	sasd...47hsd	
advertiser	2345	
usertag	123,5678,3456	Contains multi-values, ',' as segmentation.

As the linear model may not be the best model, you are asked to experiment further in order to find a more optimal model. The model you developed can be tested over the validation set. and you should explain your approach and discuss your results in the individual report. This does include, but is not limited to, reporting the performance metrics as defined in Section 3.3 of your solution and the parameter turning results on the validation set.

Note that your bidding strategy described here may be part of the more complex model (either "cleverly" combined one or further developed one based on multiagent reinforcement learning) for the group. Therefore, it is advised, although not mandatory, that you try different models as a group in this section.

Some directions: A non-linear bidding strategy (e.g. ORTB) [2], a lift-based bidding [3] or using other CTR estimators. A collection of research papers can be found here.

3.2.5 *Problem 5: Multiagent Bidding Strategy (Submission)*

- 20/50 marks

Develop a game theory (multiagent) based approach for the bidding (you might want to make use of the basic bidding algorithms developed in problem 4 to start with against the submissions from other groups. Hint 1: Game theory is useful for taking into the competition among different group submissions. See the definition of Winning criterion #2 in Sec 3.3. Next, pick up the best bid strategy, and make your bids on the given test set.

3.3 Evaluation

Alongside the final report, you are required to submit the bid prices for the testing set, generated using your group's best performing model from problem 5. Performance will be evaluated based on the following metrics, within a limited budget of $6,250$ CNY fen.

- Clicks (Num. of Clicks from Winning Impressions) **the main metric to rank the submissions**
- Click-Through Rate (Num. of Clicks over Winning Impressions)
- Spend (Total Money Paid) **spend all the budget to be a valid submission**
- Average CPM (Cost Per Mille)
- Average CPC (Cost Per Click)

The bidding prices on the testing set (not the training or validation set!), should be saved as a CSV file named 'testing_bidding_price.csv' with following format:

```
bidid,bidprice
xxxxxxxxxx,13.0
yyyyyyyyyy,23.0
....
```

An example submission file will be provided with the dataset, please strictly follow the format requirements. During the evaluation, we (our script) will go through the file and only consider the bids until the accumulated cost is higher than or equal to the allowed budget $6,250$ CNY fen.

We would expect you to produce two final submissions (therefore files) to account for two different ways of calculating the winning and the actual paid price. Winning criterion #1: the winning is determined if $bid \geq payprice$ and the actual paid price is the $payprice$ from the data. Winning criterion #2: the winning is determined if $bid \geq payprice\&otherSubmittedBids$, and pay the highest among $payprice\&otherSubmittedBids$. Note that the latter setting is still the second price auction, but accounts for the competition among the bids coming from different groups as specified by Problem 5.

3.4 Format of your reports

Your report should consist of the following sections: 1) the introduction section briefly explains the problem context and what you intend to do and summarises your results; 2) the related work section briefly summarises the related work of the specific problems you have found from the research literature. You could use https://scholar.google.co.uk/ or make use of the list here to obtain the related research papers; 3) the approach and result section gives your answers to each of the specific questions and provide the evaluation results and your discussions. 4) the conclusion section concludes your report and point out the potential direction to improve your report (e.g., if you have time, you will do...).

4. REFERENCES

[1] Weinan Zhang, Shuai Yuan, Jun Wang, and Xuehua Shen. Real-time bidding benchmarking with ipinyou dataset. *arXiv preprint arXiv:1407.7073*, 2014.

[2] Weinan Zhang, Shuai Yuan, and Jun Wang. Optimal real-time bidding for display advertising. In *Proceedings of the 20th ACM SIGKDD international conference on Knowledge discovery and data mining*, pages 1077–1086. ACM, 2014.

[3] Jian Xu, Xuhui Shao, Jianjie Ma, Kuang-chih Lee, Hang Qi, and Quan Lu. Lift-based bidding in ad selection. In *AAAI*, pages 651–657, 2016.