Modeling Advertiser Bidding Behaviors in Google Sponsored Search with a Mirror Attention Mechanism

Suqi Liu*
Princeton University
suqi@princeton.edu

Liang Liu
Google
liangliu@google.com

Sugato Basu
Google
sugato@google.com

Jean-François Crespo
Google
jfcrespo@google.com

ABSTRACT

In this paper, we present a data-driven approach to predicting advertiser's bid prices in a sponsored search system, like Google Search Ads, with a new type of attention-based sequence learning model. Instead of characterizing the advertisers' bidding behaviors with explicit assumptions (e.g., rationality models), as done in previous work [15], we treat their bid adjustments as response to observable metrics (e.g., impression count, click-through rate) and directly predict the bid prices using recurrent neural networks combined with a novel attention mechanism. The proposed model consists of two recurrent neural networks, for capturing the dynamics of metric sequence and bid sequence respectively, connected by a mirror attention layer formulation that transfers location information from metrics to bids. We evaluate the performance of the proposed model, along with other baselines, on advertiser bidding history data extracted from Google Search Ads system logs. We also demonstrate the generality of the new mechanism by experimenting on another domain: air quality prediction. Our empirical results show the effectiveness of the modeling approach and the new mechanism — we see a significant improvement over the baseline models for both advertiser bid prediction and air quality prediction tasks.

KEYWORDS

Sponsored Search; Advertiser Modeling; Bid Prediction; Sequence Learning; Attention Model

ACM Reference Format:
Suqi Liu, Liang Liu, Sugato Basu, and Jean-François Crespo. 2019. Modeling Advertiser Bidding Behaviors in Google Sponsored Search with a Mirror Attention Mechanism. In *Proceedings of AdKDD (AdKDD' 19).* ACM, New York, NY, USA, 6 pages.

1 INTRODUCTION

Search advertising has become a major source of monetization on the Internet, and plays a critical role in commercial search engines. Understanding advertisers' behaviors is essential for optimizing the auction system underlying the sponsored search (e.g., tuning the reserved price) and improving the efficiency of the associated ad market. In a sponsored search system like Google Ads, an advertiser creates advertising campaigns targeting different keyword queries. Once a specific query is issued by a search system user, the relevant campaigns are chosen to participate in an ad auction. Generalized second-price (GSP) and its variants are among the more commonly used auction mechanisms in sponsored search to select the relevant ads. The bid price (named maximum cost-per-click bid in Google Ads), also referred to simply as the bid, is the maximum amount of money an advertiser is willing to pay for an actual click. It is one of the important values an advertiser controls in a campaign. This value together with quality of the ad determines the auction score of the ad that enters the auction, which finally clinches the ranking of the ad (the ad will not be shown if a particular reserve score is not met).

There can be various reasons why advertisers make changes to their bids. Some do not concern the sponsored search and are just external causes, e.g., sales performance of a product. While excluding those outside information, the bid changes would be due to the advertisers' response to the observable metrics in the system such as impressions, click-through rate, etc. In particular, as more and more advertisers adopt auto bidding strategies, the bidding behaviors become increasingly relying on the metrics. Even if the bids are assumed to be merely the consequences of metrics, predicting them still poses many difficulties. First, the ads usually come from a large variety of domains — bids related to luxury products such as a diamond ring can be largely different from those of regular consumer products such as packs of gum. They even vary a lot within the same business category, e.g., budget wine and fine wine. Second, advertisers have very different goals towards advertising. Many advertisers hope viewers will click their ad — but that's not always the main goal. Some are primarily interested in brand exposure, and are happy to have people to see the ad, even if they don't click it. Others optimize for high ranks in the list of advertisements. Moreover, the advertisers may not always bid rationally, leading to bidding patterns that are not optimal for their goals. Finally, there can be large volatility in the bids. They even vary a lot within a short time period. For all these reasons and more, predicting bidding behaviors is quite challenging.

Prior work on the subject mostly focuses on modeling advertiser's bid strategy with explicit assumptions [1, 15]. The payoffs and interactions among all competing advertisers are discussed in depth with full or limited rationality. This approach can have a number of drawbacks. First, the utility function needs to be close enough to reality. In other words, the value of a click for an advertiser has to be accurate. Second, modeling the interactions among all competitors is not scalable and quickly becomes intractable as the number of participating campaigns grows. Third, such models generally ignore the different advertising goals, which depend on numerous metrics or their combinations. For the reasons above, the game theoretical modeling is seldom used in practice.

*This work was done while the author was at Google.

AdKDD' 19, August 5, 2019, Anchorage, Alaska, USA

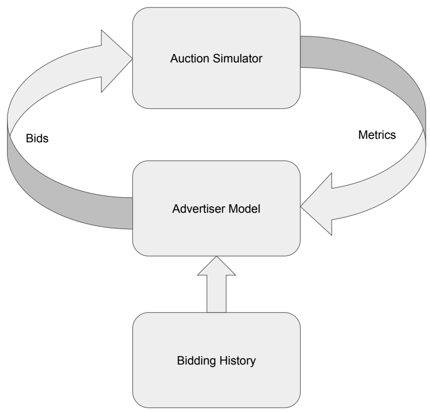

Figure 1: Ad simulation system

In this work, we take a data-driven approach to modeling advertiser bidding behaviors — it learns a bid response model of advertisers from bidding history without strong explicit assumptions like rationality. We think of the advertiser's bid changes as response to metrics, which are available to the advertiser as feedback from the sponsored search auction system, while respecting the bidding patterns. There are two types of dynamics in this modeling approach: (a) the evolution of metrics driven by the machinery of the environment (auction system), and (b) the progression of bids based on the actions taken by the agents (advertisers). Under this setting, it is useful to consider advertiser models that capture both these aspects of the dynamics. We start with a simple advertiser model that assume no response, and then experiment with successively more complex and realistic models. When we discuss the sequence models, the metric sequence is called the "driver" sequence and the bid sequence is called the "response" sequence. The single sequence model deploys a recurrent neural networks (RNN) on the driver sequence alone, while in the double sequence model, a separate network is used for each of them. We also apply the conventional attention mechanism to the driver sequence and further develop a novel mechanism termed "mirror attention", which transfers the similarity information from the states of the driver sequence to those of the response sequence.

Once the advertiser model is built, it can be used to study the individual advertiser behavior or connected to an auction simulator to give insights into ways to analyze the market and effectively improve the auction system. Accordingly, we design an ad simulation system as shown in Figure 1. The auction simulator simulates the GSP auction based on input bids from advertiser model and the historical data. The output of the auction simulator is aggregated performance metrics, which are the same as those displayed to the advertisers in the real system. The advertiser model then takes as input the bid and metric history, and outputs new bids.

To assess the performance of the advertiser model, we collected advertiser bidding history data for a particular time period in 2018 from Google Search Ads. We also evaluate the proposed attention mechanism on a publicly-available air quality dataset.

2 RELATED WORK

Modeling advertiser bidding behaviors in sponsored search has attracted increased attention from the research community [2, 14–16]. Earlier work focused on the game theoretic modeling approach, as we discussed in Section 1. Recent work [3, 7, 11, 13] has employed the data-driven approach. Broder et al. [3] and Ren et al. [13] considered a simple functional relationship between bid and features, while Cui et al. [7] predicted the bid landscape by combining templates and decision trees. These approaches do not model time-dependant dynamics. On the other hand, Long et al. [11] modeled the bids as a time series, however, the dependency on metrics was left out.

The prediction task we discuss here is also related to classic time series analysis[4, 12], with modern emphasis on learning models [6, 9]. The major difference is that our model structures involve the driver and response sequences and address the response relation between them.

3 ADVERTISER MODELING

Some widely adopted advertiser models take the path of treating the sponsored search auction as a multi-player game [1, 5]. By defining an appropriate utility function, the optimal bids are given by the equilibrium of the game. The game is usually solved by some specific algorithm together with an auction simulator. Due to the fact that optimal response is not observed in practice, some models take into account the willingness, capability and constraints of advertisers to model their imperfect rationality in setting bids [15]. This has led to a body of work in advertiser modeling that focuses on capturing the rational and irrational parts of bidding behaviors explicitly through specific model assumptions [14, 15].

We start with a different view for modeling advertiser bidding behaviors. To begin with, a brief overview of the Google Ads system is provided here. Most of the definitions are derived directly from Google Ads support page.[1] Registered customers of Google Ads can create advertisement campaigns. A campaign may contain multiple ad groups, targeting different search query terms (known as keywords). The advertiser can set a base bid for the campaign and bid multipliers for each ad group. The final bid for each query is the product of the base bid and its multiplier for the ad group. The advertiser controls two values in each campaign:

- **Maximum CPC bid** – a number to determine the highest amount that the advertiser is willing to pay for a click on the ad
- **Daily budget** – an amount to specify how much, on average, the advertiser would like to spend each day

The advertiser is also able to observe certain performance metrics for their campaign. Some of these metrics are as follows:

- **Impressions** – a count of the number of times the ad is shown on a search result page or other sites on the Google Network
- **Clicks** – a count of the number of times the users clicked on the ad

[1] https://support.google.com/google-ads/, snapshot on 2018.12.07

- **Conversions** – a count of the number of actions when someone interacts with the ad and then takes an action that is defined as valuable to advertiser's business
- **Ad Rank** – a value that is used to determine the ad's position (where ads are shown on a page relative to other ads) and whether the ad will show at all
- **Actual CPC** – the final amount that the advertiser is charged for a click

In addition to the raw metrics above, some other metrics are derived by computing the following ratios:

- **Click-through rate (CTR)** – a ratio showing how often people who see the ad end up clicking it
- **Conversion rate** – the average number of conversions per ad click, shown as a percentage

The first step of our data-driven approach is to discretize the data. Each day is chosen to be one time-step, so all metrics are aggregated on a daily basis. We also consider the advertiser model on the campaign level. In each step, the target is to predict the bid in the next step given all the bid and metric history. Both bids and metrics are normalized for each campaign to resolve the heterogeneity. Let $\mathbf{x}_t \in [0,1]^m$ denote the input vector of metrics at time t, where m is the input dimension, and let $\hat{\mathbf{y}}_t \in [0,1]$ be the output from the advertiser model, which is the predicted response bid at time t. Then, we consider the following functional form:

$$\hat{\mathbf{y}}_{t+1} = f(\mathbf{x}_0, \ldots, \mathbf{x}_t),$$

where f denotes generally the sequential prediction model. We incorporate the true bid $\mathbf{y}_t$ at time t as part of $\mathbf{x}_t$.

In addition to metrics, some demographic information of the advertisers can also be retrieved — it is reasonable to believe that this facilitates the understanding of their behaviors as well. The following features, which are usually reported by the advertiser, are also used in the prediction model:

- **Language** – the primary language of the ads landing page
- **Country** – the targeted country of ads
- **Vertical** – the specific markets that the ads are targeting
- **Classification** – the commercial sectors, business categories and business attributes of the advertiser

Let $\mathbf{z}$ denote the features corresponding to these attributes of the advertiser. The prediction model can now be written as:

$$\hat{\mathbf{y}}_{t+1} = f(\mathbf{x}_0, \ldots, \mathbf{x}_t; \mathbf{z}).$$

The parameters of the sequential prediction model are estimated on advertiser's bidding history by minimizing an appropriate loss between predicted bid $\hat{\mathbf{y}}_{t+1}$ and the target, i.e., the true bid at the next step $\mathbf{y}_{t+1}$. Suppose we have N sequences each of length T. The training phase becomes solving the following optimization problem:

$$\min_{\boldsymbol{\theta}} \sum_{i=1}^{N} \sum_{t=1}^{T} \left(\hat{\mathbf{y}}_t^{(i)} - \mathbf{y}_t^{(i)}\right)^2,$$

where $\boldsymbol{\theta}$ denotes in general the model parameters. Here, the squared loss is used. By modeling the bidding behaviors as response to observable metrics, the advertiser's bidding strategies are implicitly learned by the model that is trained from historical data.

4 SEQUENTIAL PREDICTION MODELS

In this section, we describe several baseline models and the proposed mirror attention mechanism. The models are introduced in the order of increasing complexity, encapsulating intuitions derived from advertiser modeling. For all the models, each type of the advertiser features is first passed through an embedding layer and then served as input to the prediction layer.

4.1 Naïve forecast

A simple prediction model [12] assumes no response to metrics and merely outputs the actual bid from the last step. It is sometimes referred to as "naïve forecast" in learning community. The predicted bid at time $t+1$ is simply

$$\hat{\mathbf{y}}_{t+1} = \mathbf{y}_t.$$

4.2 Linear regression

The linear regression model [12] makes predictions based on the inputs within a time window of certain size, while the output has a linear dependence over all inputs. Considering the metrics in s time steps, the model prediction takes the following form (as shown in Figure 2):

$$\hat{\mathbf{y}}_{t+1} = \sum_{i=1}^{s} \boldsymbol{\beta}_i^\top \mathbf{x}_{t-i+1},$$

where $\boldsymbol{\beta}_i \in \mathbb{R}^m$ is the weight vector for inputs at time $t-i+1$. The linear regression depending solely on previous targets is also called auto-regressive (AR) model in statistics.

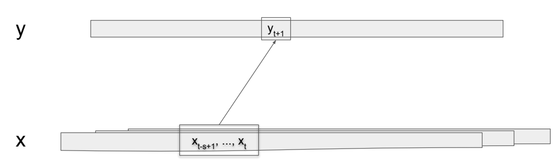

Figure 2: Regression model

4.3 Single sequence

The sequence modeling approach in machine learning extends the linear regression model by allowing nonlinear and long-range dependencies. Practically, the inputs at all previous time steps are fed sequentially into a recurrent neural network (RNN) such as long-short term memory (LSTM) [8]. A fully-connected layer is stacked on top of the outputs of the RNN (shown in Figure 3). The prediction of the model is emitted from the fully-connected layer. Let $\mathbf{h}_t$ be the hidden state vector of the RNN. The update rule can be written generally as

$$\mathbf{h}_t = f(\mathbf{x}_t, \mathbf{h}_{t-1}).$$

In a regression model, the output only depends on part of inputs within the fixed window. Long range dependencies are not well captured. The recurrent structure allows the information over a relatively longer time period to be still relevant. This is essential in bid prediction since the response can be the result of some accumulating factors.

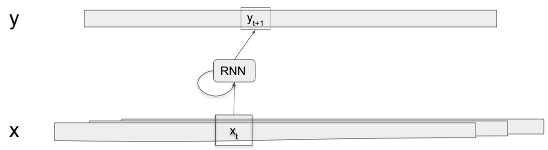

Figure 3: Single sequence model

4.4 Causal attention

Causal attention is a local attention mechanism that preserves the causality over the series. At each time step t, the hidden states within a window of size s from $t-s$ to $t-1$ is used to compute a context vector $\mathbf{c}_t$ with a score function. For two vectors $\mathbf{u}$ and $\mathbf{v}$, three score functions are commonly used:

$$score(\mathbf{u}, \mathbf{v}) = \begin{cases} \mathbf{u}^\top \mathbf{v} & dot \\ \mathbf{u}^\top W \mathbf{v} & general \\ \mathbf{w}^\top \tanh(W[\mathbf{u}; \mathbf{v}]) & concat \end{cases},$$

where the vector $\mathbf{w}$ and the matrix W are model parameters to be trained. The attention weight vector is then computed by

$$\boldsymbol{\alpha}_t = softmax(score(H_t, h_t)),$$

where $H_t = [\mathbf{h}_{t-s+1}, \ldots, \mathbf{h}_{t-1}]$ is the matrix where the columns are the hidden states within the time window, and the score function is applied to the matrix column-wise, while the softmax function converts the scores into a discrete probability distribution. Finally, the context vector is calculated as the expectation of the hidden states over the distribution $\boldsymbol{\alpha}_t$:

$$\mathbf{c}_t = H_t \boldsymbol{\alpha}_t.$$

In our formulation, we use the attention mechanism over an LSTM layer. The output of the model is built on top of the RNN and the context vector $\mathbf{c}_t$ by a fully-connected layer (see Figure 4).

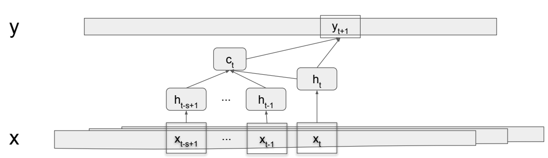

Figure 4: Causal attention model

The attention mechanism facilitates the prediction at step t by querying relevant locations within the time window. This allows long range experience to be directly reusable as reference.

4.5 Double sequence

In addition to modeling the dependency of $\mathbf{y}_{t+1}$ on the history of $\mathbf{x}_0, \ldots, \mathbf{x}_t$, a double sequence model takes into account the separate evolution of $\mathbf{y}_i$'s. Alongside the RNN over the input sequence, another RNN that takes $\mathbf{y}_0, \ldots, \mathbf{y}_t$ as inputs is used to capture the dynamics of the targets. Similar to previous models, a fully-connected layer is stacked on top of the outputs of the RNNs. The final output of the model is a combination of the hidden states encoding the input history and those encoding the target history (see Figure 5).

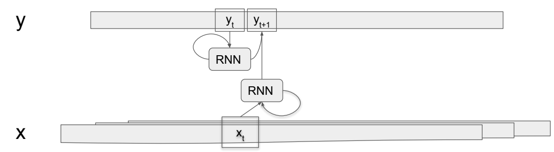

Figure 5: Double sequence model

This architecture is a special case of single sequence that forces the dynamics of the outputs to be modeled separately. In the sequential prediction context, the isolation of inputs and outputs becomes natural — the inputs evolves independently according to the environment, while the output sequence can have separate dynamics in addition to the response to inputs. For example, the bid sequence may satisfy some smoothness assumption that does not allow abrupt change (e.g., based on advertiser policy).

4.6 Mirror attention

This novel formulation extends the double sequence with a specialized attention mechanism. In attention-based learning, the attention scores indicate where to focus in the sequence. Similar to causal attention, the attention weight vector is computed based on the hidden states of the driver sequence. But instead of using this similarity information with the driver sequence, it is applied to the response sequence to aggregate the target history into a context vector. Let G_t be the matrix of output hidden states within a window s in the response sequence. The context vector in mirror attention is defined as

$$\mathbf{c}_t = G_t \boldsymbol{\alpha}_t.$$

The attention weight vector $\boldsymbol{\alpha}_t$ is computed with the states of the driver sequence as before. Figure 6 shows this mechanism graphically.

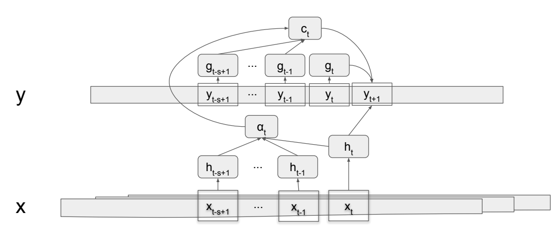

Figure 6: Mirror attention model

If attention mechanism is understood as using historical information to facilitate prediction, it is reasonable to combine the information not only from the environment but also from the actions taken before. When hard attention is used, the mirror attention mechanism is equivalent to querying key-value pairs composed of the driver and response states.

5 ADVERTISER DATA

The advertiser data consist of two parts: (a) bidding history with metrics, and (b) information regarding the campaign, as discussed in the modeling approach (Section 3). The advertiser bidding history dataset is fully anonymized — all identifying information is removed to protect the privacy of the advertisers. Also, for privacy concerns, we are not able to release the experiment data in this study for now.

5.1 Bidding history

The dataset is collected from auction system logs during a particular period in 2018 in Google Search Ads. We retrieved the average daily bid along with all metrics that are displayed to an advertiser in the Google Ads dashboard. We noticed that some small campaigns were created for testing purposes and their bidding behaviors are largely experimental. So campaigns with number of clicks smaller than a threshold were removed. Campaigns that are not active during this whole period were also excluded.

The raw metrics usually exhibit large variations across campaigns. Therefore, we normalized all raw metrics using a min-max normalization for each campaign to alleviate this problem. The normalized bids and metrics, fall into a range from 0 to 1. Hence, the bid we are predicting is a relative value with respect to the minimum and maximum bid of each campaign.

In addition to the raw metrics, we also included several derived metrics, which are the ratios of raw metrics. The following derived metrics are used:

- **Click-through rate (CTR)** – # clicks / # impressions
- **Conversion rate (CVR)** – # conversions / # clicks
- **Cost-budget ratio (CBR)** – cost / budget
- **Cost per click (CPC)** – cost / # clicks

Since some commonly seen bidding strategies are based on those ratios, it is natural to believe that they have direct impacts on the bids or at least are useful for prediction.

5.2 Campaign information

In addition to the logs, we also obtained background information of each campaign as features, listed in Table 1 with their dimensions. The features are treated as categorical inputs into the model. As the categories usually have a long tail distribution, i.e., large numbers of categories contain only a few instances, we merged all small categories into a single one. Each feature is connected to the model using an embedding layer with a specific embedding dimension.

Table 1: Campaign background information with their dimensions

Feature	Dimension
Language	91
Country	225
Vertical	1,426
Classification	5,006

6 EXPERIMENTS

We evaluate the models on both the air quality dataset and advertiser dataset. The models in our experiments will be referred to by their abbreviations – NAIV: naïve forecast; REGR: regression; SSEQ: single sequence; CATT: causal attention; DSEQ: double sequence; MATT: mirror attention.

6.1 Air quality

The air quality dataset [10] contains the hourly PM2.5 data of the US Embassy in Beijing, together with meteorological data from Beijing Capital International Airport, from Jan 1st, 2010 to Dec 31st, 2014. Data from 2010 to 2013 are used for training and those in 2014 are used for testing. The dataset resembles the advertiser response in the sense that the concentration of pollutant has its own evolution but is still influenced by meteorological conditions like temperature, pressure, wind speed, cumulative hours of rain, etc. We treat the meteorological information as the driver sequence and the pollutant concentration as the response sequence.

To best mimic the serial pattern of advertiser bidding history, we chunk the long sequence into short sequences of length 3 days with 72 time steps. All features are normalized using a min-max normalization. The data are further converted to four prediction tasks: pollutant concentration after 3 hours, 6 hours, 12 hours and 24 hours, with increasing difficulty. The results are shown in Table 2.

Table 2: Comparison of model performances on air quality dataset

	3 hours	6 hours	12 hours	24 hours
NAIV	100.00	100.00	100.00	100.00
REGR	111.11	121.98	113.65	75.63
SSEQ	**87.70**	89.03	73.28	66.41
CATT	94.36	**81.44**	73.32	67.99
DSEQ	93.82	84.47	**72.72**	**65.21**
MATT	**89.89**	**81.82**	**72.82**	**65.67**

The mean squared error of baseline (naïve forecast) is normalized to be 100.00, and we present the relative values of other models' mean squared errors compared to this baseline. The smallest two prediction errors in each task are highlighted in bold. As we can see, modeling both the drive and response sequences (DSEQ or MATT) achieves better results when we increase the difficulty of the prediction task with larger predicting gap, and MATT performs consistently among the best models.

6.2 Advertiser bidding history

We sample a fraction of advertisers in Google Ads during a continuous time period. The training and testing split is according to campaigns. 80% of the campaigns are used for training and 20% for testing. We also report the relative performances of the models compared to the baseline, which is a naïve forecast model with one day time difference, i.e., simply predicting the bid using the actual advertiser bid on the last day. The results are shown in Table 3.

Evaluated on the advertiser data, the relative mean squared errors are converted to percentages same as in previous experiments,

Table 3: Comparisons of model performance on advertiser bidding history dataset

Model	Relative mean squared error
NAIV	100.00
REGR	85.62
SSEQ	80.82
CATT	74.11
DSEQ	64.96
MATT	**61.98**

where lower scores indicate better performances. Due to the complexity in advertiser bid data, especially the dynamics of the bid sequence, models such as REGR, SSEQ, CATT shows less improvement with respect to the baseline. Double sequence models such as DSEQ and MATT achieve lower errors. The best performance comes from the proposed mirror attention mechanism.

To study the impact of model complexity, we fine-tuned the parameters of MATT with several configurations of networks to see how much incremental gain it can provide. For the model parameters of MATT, we denote the length of time window in a sample as W, the dimension of hidden states in metric sequence as M, the dimension of hidden states in bid sequence as B, the score function dimension as S, and the embedding layer dimensions as a vector E. Some representative results are detailed in Table 4 as percentage improvements over the baseline.

Table 4: Parameter tuning of mirror attention model on advertiser bidding history dataset

Parameters	Relative improvement (%)
$W = 10, M = 8, B = 4, S = 10$ $E =[10, 10, 20, 20]$	22.71
$W = 30, M = 8, B = 4, S = 10$ $E = [10, 10, 20, 20]$	27.54
$W = 30, M = 16, B = 8, S = 10$ $E = [10, 10, 20, 20]$	28.25
$W = 30, M = 16, B = 8, S = 20$ $E = [10, 10, 20, 20]$	28.75
$W = 30, M = 128, B = 64, S = 40$ $E = [10, 10, 20, 20]$	**38.01**
$W = 30, M = 128, B = 64, S = 40$ $E = [10, 64, 64, 128]$	37.74

For MATT models, the prediction accuracy improves with larger model sizes. We noticed that the length of attention window W plays an important role in model performance, since it defines the look-back period for advertiser's decision-making. The dimensions of hidden states in the driver sequence and response sequence significantly contribute to the model performance as well. Score dimension and embedding layer dimension are less effective though adding extra training latency. Overall, we see substantial improvement for MATT from simple structure to complex structure on the bid prediction task, with careful parameter tuning.

7 CONCLUSION

In this paper, we introduced a new data-driven approach to advertiser bid prediction. A novel mirror attention mechanism tailored to the sequential prediction task was proposed. We experimented with various models on the advertiser bidding history data, as well as a benchmark air quality dataset, and showed the advantage of the proposed mechanism. This work is a first step in our attempts towards understanding advertiser behaviors via sequence modeling – this study shows the potential of the proposed model that enables us to simultaneously learn the advertisers' response and optimize the system for better performance. In future work, we will continue working on more complex model structures to further improve the prediction power. For the proposed mirror attention mechanism itself, we also see the wide applications to other prediction tasks such as user modeling, stock price prediction, etc.

REFERENCES

[1] Gagan Aggarwal, S. Muthukrishnan, Dávid Pál, and Martin Pál. 2009. General Auction Mechanism for Search Advertising. In *Proceedings of the 18th International Conference on World Wide Web (WWW '09)*. ACM, New York, NY, USA, 241–250.
[2] Kareem Amin, Michael Kearns, Peter Key, and Anton Schwaighofer. 2012. Budget Optimization for Sponsored Search: Censored Learning in MDPs. In *Proceedings of the Twenty-Eighth Conference on Uncertainty in Artificial Intelligence (UAI'12)*. AUAI Press, Arlington, Virginia, United States, 54–63.
[3] Andrei Broder, Evgeniy Gabrilovich, Vanja Josifovski, George Mavromatis, and Alex Smola. 2011. Bid Generation for Advanced Match in Sponsored Search. In *Proceedings of the Fourth ACM International Conference on Web Search and Data Mining (WSDM '11)*. ACM, New York, NY, USA, 515–524.
[4] Robert Goodell Brown. 2004. *Smoothing, forecasting and prediction of discrete time series.* Courier Corporation.
[5] Matthew Cary, Aparna Das, Ben Edelman, Ioannis Giotis, Kurtis Heimerl, Anna R. Karlin, Claire Mathieu, and Michael Schwarz. 2007. Greedy Bidding Strategies for Keyword Auctions. In *Proceedings of the 8th ACM Conference on Electronic Commerce (EC '07)*. ACM, New York, NY, USA, 262–271.
[6] Kanad Chakraborty, Kishan Mehrotra, Chilukuri K Mohan, and Sanjay Ranka. 1992. Forecasting the behavior of multivariate time series using neural networks. *Neural networks* 5, 6 (1992), 961–970.
[7] Ying Cui, Ruofei Zhang, Wei Li, and Jianchang Mao. 2011. Bid Landscape Forecasting in Online Ad Exchange Marketplace. In *Proceedings of the 17th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD '11)*. ACM, New York, NY, USA, 265–273.
[8] Sepp Hochreiter and JÃijrgen Schmidhuber. 1997. Long Short-Term Memory. *Neural Computation* 9, 8 (1997), 1735–1780.
[9] Guokun Lai, Wei-Cheng Chang, Yiming Yang, and Hanxiao Liu. 2018. Modeling Long- and Short-Term Temporal Patterns with Deep Neural Networks. In *The 41st International ACM SIGIR Conference on Research & Development in Information Retrieval (SIGIR '18)*. ACM, New York, NY, USA, 95–104.
[10] Xuan Liang, Tao Zou, Bin Guo, Shuo Li, Haozhe Zhang, Shuyi Zhang, Hui Huang, and Song Xi Chen. 2015. Assessing Beijing's PM2.5 pollution: severity, weather impact, APEC and winter heating. *Proceedings of the Royal Society of London A: Mathematical, Physical and Engineering Sciences* 471, 2182 (2015).
[11] Lili Long, Hongbin Dong, Yue Pan, Li Huangfu, Naikang Gou, and Xingmei Wang. 2015. Forecasting model for bidding behavior of advertisers based on HMM. In *2015 International Joint Conference on Neural Networks (IJCNN)*. 1–7.
[12] Douglas C Montgomery, Cheryl L Jennings, and Murat Kulahci. 2015. *Introduction to time series analysis and forecasting.* John Wiley & Sons.
[13] K. Ren, W. Zhang, K. Chang, Y. Rong, Y. Yu, and J. Wang. 2018. Bidding Machine: Learning to Bid for Directly Optimizing Profits in Display Advertising. *IEEE Transactions on Knowledge and Data Engineering* 30, 4 (April 2018), 645–659.
[14] Jiang Rong, Tao Qin, Bo An, and Tie-Yan Liu. 2016. Modeling Bounded Rationality for Sponsored Search Auctions.. In *ECAI*. 515–523.
[15] Haifeng Xu, Bin Gao, Diyi Yang, and Tie-Yan Liu. 2013. Predicting Advertiser Bidding Behaviors in Sponsored Search by Rationality Modeling. In *Proceedings of the 22Nd International Conference on World Wide Web (WWW '13)*. ACM, New York, NY, USA, 1433–1444.
[16] Yunhong Zhou, Deeparnab Chakrabarty, and Rajan Lukose. 2008. Budget constrained bidding in keyword auctions and online knapsack problems. In *International Workshop on Internet and Network Economics*. Springer, 566–576.