Bid Landscape Forecasting in Real-Time Bidding Advertising

Jiarui Qin

Apex Data and Knowledge Management Lab
Shanghai Jiao Tong University

May. 23, 2018

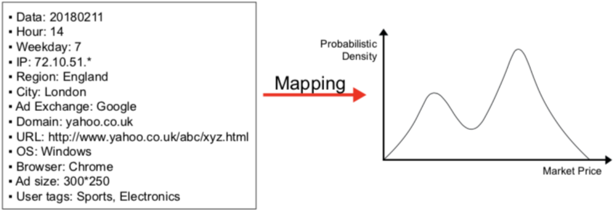
Outline

1. Problem Definition
   - Bid Landscape Forecasting
   - The Challenges
2. Parametric Approach
   - Gamma Based Distribution
   - Mixture Model
3. Non-Parametric Approach
   - Survival Tree Model
4. Our Approach: Deep Survival Analysis
   - Survival Analysis
   - Using RNN
   - Loss Function
   - Evaluation

Outline

Bid Landscape Forecasting

In RTB scenario, it is very important for DSPs(agents for advertisers) to model the market price(price for the ad impression after the auction) distribution of each bid request. The predicted market price distribution is a key part of the DSP's bidding strategy.



Auction Information

Bid Landscape

Outline

The Challenges

- Nearly impossible to model each DSPs bidding strategy.
- Censorship issue: one DSP can only know the market price of a bid request if it wins the bid. So if only use winning bids to train the model, it will introduce heavy bias.
- Market price distribution can be diverse, it is highly related to the features of the bid request.(eg.location)

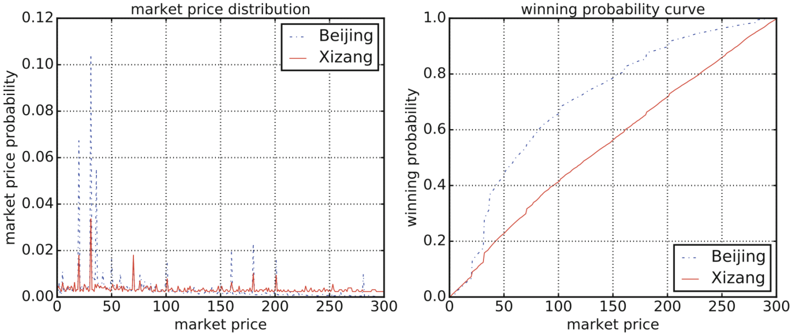
The Challenges
Censorship Issue

Average market price of winning set is much larger than that of losing set.

day	# of bids	# of win. bids	WR	EWR	WR AUC	Avg. WP	Avg. WP on W	Avg. WP on L
2013-06-06	1,821,479	1,514,416	0.83	0.83	0.89	74.86498	52.46772	185.3269
2013-06-07	1,806,062	1,524,314	0.84	0.85	0.90	72.31279	51.12051	186.9674
2013-06-08	1,634,967	1,352,038	0.83	0.83	0.87	81.14319	58.48506	189.4200
2013-06-09	1,651,630	1,366,097	0.83	0.83	0.88	81.31667	58.95707	188.2934
2013-06-10	1,920,576	1,603,798	0.84	0.83	0.91	79.83572	58.91341	185.7621
2013-06-11	1,745,905	1,461,085	0.84	0.86	0.85	79.62260	58.91626	185.8431
2013-06-12	1,657,578	1,378,728	0.83	0.85	0.84	79.99693	58.80196	184.7920

The Challenges
Diverse distribution

Different bid request has very different market price distribution.



Outline

Gamma Based Distribution

Methodology

It assumes that the market price distribution is a gamma distribution. For i-th bid record in dataset, the PDF $f_{Z_i(z_i)}$ and CDF $F_{Z_i(z_i)}$ of the market price Z_i is

$$f_{Z_i(z_i)} = \frac{1}{\Gamma(k_i)\theta^{k_i}} z_i^{k_i-1} e^{\frac{-y_i}{\theta}} \tag{1}$$

$$F_{Z_i(z_i)} = \frac{1}{\Gamma(k_i)} \gamma(k_i, \frac{b_i}{\theta}) \tag{2}$$

where z_i is market price, b_i is bidding price, Γ and γ is the gamma function and the lower incomplete gamma function. k_i and θ is the parameters of the gamma distribution of i-th record.

Accroding to gamma distribution, the predicted market price is $E[Z_i|x_i]$

$$z_i = E[Z_i|x_i] = k_i\theta = e^{bx} \tag{3}$$

Now we can derive the loss function.

$$J(b) = \frac{1}{n}\sum_i [y_i L_W^i + (1 - y_i)L_L^i] \tag{4}$$

$$L_W = ln(f_{Z_i(z_i)}) \tag{5}$$

$$L_L = ln(1 - F_{Z_i(b_i)}) \tag{6}$$

$$\hat{b} = \arg\min_b -J(b) + \alpha R(b) \tag{7}$$

where y_i is the label indicating if it is a winning record or not.

It uses a two-phase training.
First step just finds the K_i and *theta*

$$\hat{k}_i, \hat{\theta} = \arg\max J(k_i, \theta) \tag{8}$$

Second step just train b

$$\hat{b} = \arg\min_b \ \frac{1}{n} \sum_i (\hat{k}_i \hat{\theta} - e^{bx_i}) + \alpha R(b) \tag{9}$$

So, the predicted market price of x_i is e^{bx_i}.

Outline

Mixture Model

It proposes a linear regression model to predict market price z_i for i-th bid.

$$z_i = \beta^T x_i + \epsilon_i \tag{10}$$

It assumes that ϵ_i is and iid and normally distributed random variable with zero mean and σ variance. It uses negative log-likelyhood as loss function.

$$\sum_{i \in W} -\log(\phi(\frac{z_i - \beta_{lm}^T x_i}{\sigma})) \tag{11}$$

W is the winning set.

We need to consider censorship issue, and take the losing bid records into consideration.

$$P(z_i < b_i) = P(\epsilon_i < b_i - \beta_{clm}^T x_i) = \Phi(\frac{b_i - \beta_{clm}^T x_i}{\sigma}) \tag{12}$$

$$\sum_{i \in W} -\log(\phi(\frac{z_i - \beta_{clm}^T x_i}{\sigma})) + \sum_{i \in L} -\log(1 - \Phi(\frac{b_i - \beta_{clm}^T x_i}{\sigma})) \tag{13}$$

where L is losing set. Minimize the above loss.

β_{lm} and β_{clm} both have bias. Needs to mix them using winning rate.

$$z_i = [P(b_i > z_i)\beta_{lm} + (1 - P(b_i > z_i))\beta_{clm}]^T x_i = \beta_{mix}^T x_i \tag{14}$$

Using logistic function to predict winning rate.

$$P(z_i < b_i) = \frac{1}{1 + e^{-\beta_{lr}^T x_i}} \tag{15}$$

If we use Eq. 12 to calculate winning rate, it may introduce hidden dependency.

Outline

It uses Kaplan-Meier Product-Limit method to model the censored data. According to survival theory, the probability of losing an auction with bidding price b_x is

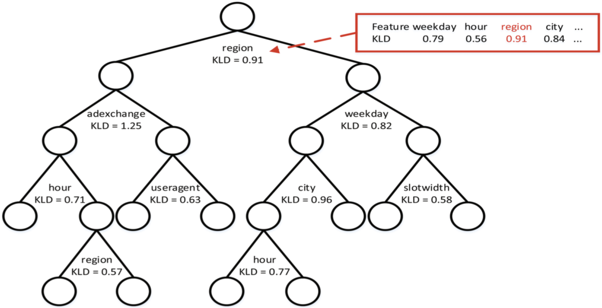
$$l(b_x) = \prod_{b_j \le b_x} \frac{n_j - d_j}{n_j} \tag{16}$$

where n_j is the number of auctions that cannot be won with bidding price $b_j - 1$, and d_j is the number of auctions whose market price is just $b_j - 1$. So, the winning probability with bidding price b_x is $w(b_x) = 1 - l(b_x)$, and market price probability is $p(z) = w(z+1) - w(z)$.

Survival Tree Model

Methodology: Binary Decision Tree

As the market price distribution is highly related to the features of the bid request, it uses binary decision tree to classify the bid requests and all the bid requests that are in same leaf node have same market price distribution which can be calculated using Kaplan-Meier Product-Limit method.



Algorithm 1. K-Means clustering with KL-Divergence

Input: Training sample $S = \{s_1, s_2, ..., s_n\}$; Attribute A_j;
Output: KL-Divergence D^j_{KL} over attribute A_j, Data clusters S^1 and S^2;

```
Randomly split the data into two parts S^1 and S^2;
while not converged do
  E-step:
  Get price probability distribution Q_1 for S^1 and Q_2 for S^2;
  M-step:
  for all M_k, k ∈ {1, 2, 3, ..., n} do
    Calculate the K_1 between M_k and Q_1 by Eq. (4);
    Calculate the K_2 between M_k and Q_2 by Eq. (4);
    Update S^1 or S^2 by comparing with K_1 and K_2;
  end for
  Calculate the D^j_KL between Q_1 and Q_2 by Eq. (4);
end while
Return D^j_KL, S^1 and S^2;
```

s_i is the set of training samples with the same value for attribute A_j. For $s_1, s_2, ..., s_n$, we have n corresponding market price probability distributions $M_1, M_2, ..., M_n$

Survival Tree Model

Methodology: Binary Decision Tree

Algorithm 2. Building Decision Tree with K-Means clustering

Input: Training sample S which contain N attributes;

1: **for** all attribute $A_j, j \in \{1, 2, 3, ..., N\}$ **do**
2:     Calculate the KL-Divergence D_{KL}^{j} for attribute A_j by Algorithm 1;
3: **end for**
4: $D_{\mathrm{KL}}^{\mathrm{best}} = \max\{D_{\mathrm{KL}}^{1}, D_{\mathrm{KL}}^{2}, ..., D_{\mathrm{KL}}^{j}, D_{\mathrm{KL}}^{N}\}$;
5: Find A_{best} with $D_{\mathrm{KL}}^{\mathrm{best}}$;
6: Create a decision node that splits on A_{best};
7: Split the decision node into two nodes S^1 and S^2;
8: Return new nodes as children of the parent node

Run the building process recursively, and stop until the length of sample data in node is less than a predifined value. When a new bid request comes, classify it to some leaf node, and get the market price distribution.

Outline

Survival Analysis

We define a instant winning rate as

$$h(b) = \lim_{\Delta b \to 0} \frac{Pr(b \leq z < b + \Delta b | z \geq b)}{\Delta b} \tag{17}$$

it is like $\frac{d_j}{n_j}$ in Eq.16. And we can drive that

$$\begin{aligned} h(b) &= \lim_{\Delta b \to 0} \frac{Pr(b \leq z < b + \Delta b)/Pr(z \geq b)}{\Delta b} \\ &= \lim_{\Delta b \to 0} \frac{S(b) - S(b + \Delta b)/Pr(z \geq b)}{\Delta b} \\ &= -\frac{S'(b)}{S(b)} = 1 - \frac{S(b+1)}{S(b)} \end{aligned}$$

where $S(b) = Pr(z \geq b)$, it is losing probability of bidding price b.

Survival Analysis

So, losing probability at bidding price b is $S(b)$

$$S(b) = \prod_{i<b}(1 - h(i)) \tag{18}$$

and winning probability $W(b)$ is

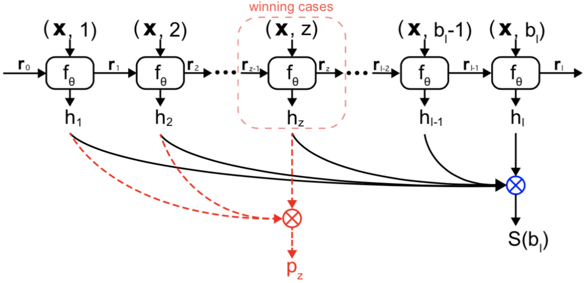
$$W(b) = 1 - S(b) \tag{19}$$

and market price probability $p(z)$ is

$$p(z) = h(z)S(z) \tag{20}$$

Outline

We propose a RNN model to fit the survival model. the l th RNN cell predicts the instant winning rate h_l given the bid request feature x_i and a bid price b based on the previous events.



Outline

Loss Function

For all bid records, we can utilize a cross entropy loss, because we can treat it as a binary classification problem.
For all winning records, we use negative log-likelyhood to discribe how good is the model. We combine the two loss functions.

$$L_{censor} = -\sum_{x_i, b_i \in D} c_i \log S(b_i|x_i) + (1 - c_i)\log(1 - S(b_i|x_i)) \tag{21}$$

$$L_z = -\sum_{x_i, z_i \in D_{win}} \log(p(z)) \tag{22}$$

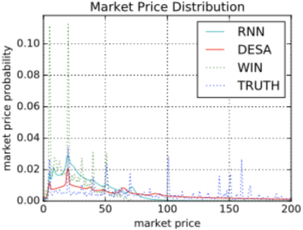
$$L = \alpha L_z + (1 - \alpha) L_{censor} \tag{23}$$

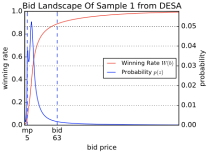
where c_i is the label censor(losing) or not. Just minimize L.

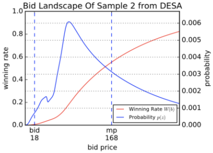
Outline

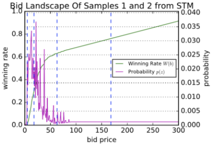
Evaluation

Three metrics: AUC, Log-Loss, ANLP(Average Negative Log Probability)









Reference

- Wush Chi-Hsuan Wu, Mi-Yen Yeh, and Ming-Syan Chen. 2015. Predicting Win- ning Price in Real Time Bidding with Censored Data. In KDD.
- YuchenWang,KanRen,WeinanZhang,JunWang,andYongYu.2016. Functional bid landscape forecasting for display advertising. In ECML-PKDD.
- W.Y.Zhu,W.Y.Shih,Y.H.Lee,W.C.Peng,andJ.L.Huang.2017. A gamma-based regression for winning price estimation in real-time bidding advertising. In 2017 IEEE International Conference on Big Data (Big Data). 16101619.