Collaborative filtering applied to real-time bidding

Pedro Chahuara

XRCE

May 2016

Summary

Context : real-time reserve price prediction

Revenue maximization through high-dimensional sparse linear regression

Strategies to deal with lost auctions

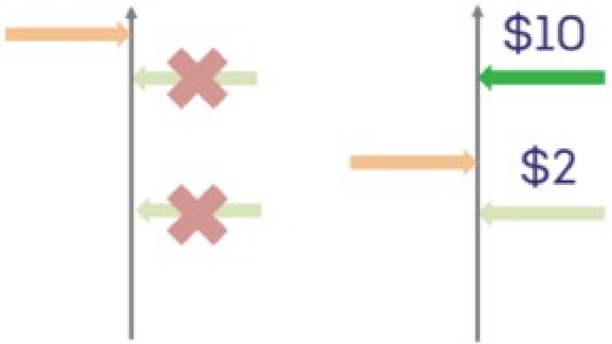
Aalen Aditive model : Probability distribution estimation with censored data

Adaptive model for revenue maximization

Results

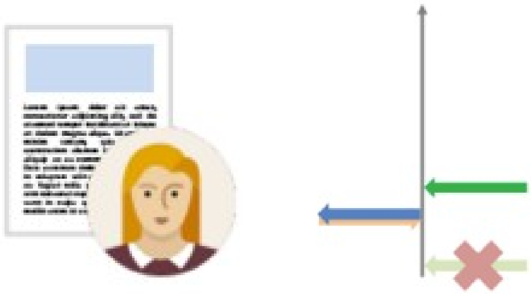
Context : real-time reserve price prediction

2nd price auction mechanism



- Buyers send bids
- Publisher sets a reserve price (the auction won't be sold below)
- Buyer with the highest bid wins the auction and displays its ad, paying the maximum between the second highest bid and the reserve price

AlephD provides its clients real-time reserve price prediction engine

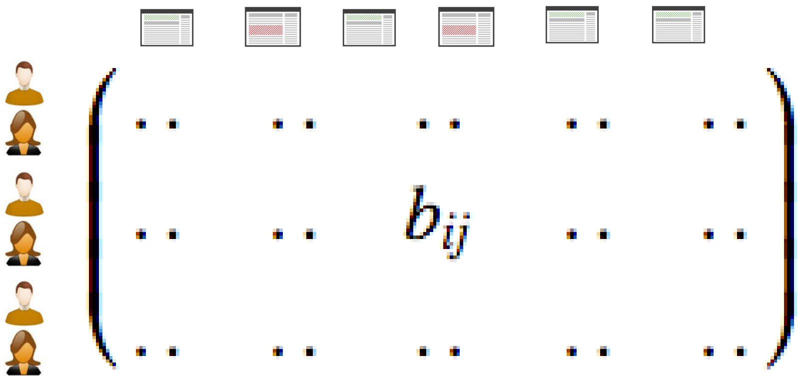


- Before each auction, we know the user and the ad placement (tag)

Context : real-time reserve price prediction

Collaborative filtering methods may be used to build user-specific and tag-specific features

- Bid prediction is obtained by combining user features and tag features



Matrix factorization using Alternating Least Squares

- X and Y are vectors of latent variables
- Discretized floor price f ϵ [f1...fk]
- The bid is the sum of user bias, tag bias and cross terms

$$R^{(k)} = \beta^{(k)} + (X_u^{(k)})' Y_p^{(k)} + \epsilon^{(k)}$$

- Loss function
- Observations are temporally weighted
- Generalized Tikhonov regularization

$$L(\hat{\beta}^{(k)}, \hat{X}^{(k)}, \hat{Y}^{(k)}) = \sum_{a \in \mathcal{D}_T} \gamma^{(T-t_a)} (R_a^{(k)} - \hat{\beta}^{(k)} - (\hat{X}_{u_a}^{(k)})' \hat{Y}_{p_a}^{(k)})^2$$
$$+ \sum_{u \in \mathbb{U}} \|\hat{X_u}^{(k)} - X_0^{(k)}\|_\Omega^2 + \sum_{p \in \mathbb{P}} \|\hat{Y_p}^{(k)} - Y_0^{(k)}\|_\Gamma^2 + \|\hat{\beta}^{(k)} - \beta_0^{(k)}\|_\Sigma^2$$

On-line Matrix factorization

- After observing an auction a, the parameters of the model are estimated iterating the following equations:

$$\hat{X}_{u_a} = X_0 + \left(\gamma^{\Delta t_{u_a}} X_{u_a,cov} + \hat{Y}_{p_a}\hat{Y}'_{p_a} + \Omega^{-1}\right)^{-1} \left(\gamma^{\Delta t_{u_a}} X_{u_a,obs} + (R_a - \hat{\beta} - X'_0\hat{Y}_{p_a})\hat{Y}_{p_a}\right)$$

$$\hat{Y}_{p_a} = Y_0 + \left(\gamma^{\Delta t_{p_a}} Y_{p_a,cov} + \hat{X}_{u_a}\hat{X}'_{u_a} + \Gamma^{-1}\right)^{-1} \left(\gamma^{\Delta t_{p_a}} Y_{p_a,obs} + (R_a - \hat{\beta} - Y'_0\hat{X}_{u_a})\hat{X}_{u_a}\right)$$

$$\hat{\beta} = \beta_0 + \left(\gamma^{\Delta t}\beta_{cov} + 1 + \Sigma^{-1}\right)^{-1} \left(\gamma^{\Delta t}\beta_{obs} + (R_a - \hat{X}'_{u_a}\hat{Y}_{p_a})\right)$$

- The following updates take place :

$$X_{u_a,cov} \leftarrow \gamma^{\Delta t_{u_a}} X_{u_a,cov} + \hat{Y}_{p_a}\hat{Y}'_{p_a}$$

$$X_{u_a,obs} \leftarrow \gamma^{\Delta t_{u_a}} X_{u_a,obs} + (R_a^{(k)} - \hat{\beta} - X'_0\hat{Y}_{p_a})\hat{Y}_{p_a}$$

$$Y_{p_a,cov} \leftarrow \gamma^{\Delta t_{p_a}} \times Y_{p_a,cov} + \hat{X}_{u_a}\hat{X}'_{u_a}$$

$$Y_{p_a,obs} \leftarrow \gamma^{\Delta t_{p_a}} \times Y_{p_a,obs} + (R_a^{(k)} - \hat{\beta} - Y'_0\hat{X}_{u_a})\hat{X}_{u_a}$$

$$\beta_{cov} \leftarrow \gamma^{\Delta t}\beta_{cov} + 1$$

$$\beta_{obs} \leftarrow \gamma^{\Delta t}\beta_{obs} + (R_a - \hat{X}'_{u_a}\hat{Y}_{p_a})$$

Lost Autions in the On-line Setting

- Two cases
- floor>B1 : Full censorship
- floor>B2 : Half censorship

- **How to update the parameters of the model ?**
- **Possible Strategies :**

  - Ignore those auctions : It does not perform exploration
  - Set the censored variable to 0.
  - Set the censored variable to the treshold value(floor).
  - Keep a distribution of the variables and try to estimate the expectation

Aalen Additive Model

- Linear regression model to estimate the hazard function with covariates $x_{1..T}$:

$$\lambda(t) = b_0(t) + b_1(t)x_1 + \ldots + b_N(t)x_T$$

- One model for each (discrete) value of t
- The cumulative distribution function can be obtained as :

$$\Phi(t) = 1 - \sum_{z=0}^{t} \lambda(z)$$

- Goal : To keep two independent distributions : B1 and B2

Aalen Additive Model - Learning

- Training Set, T : tuples (x1,..xT,**t**) where $t \in \{0, 1, ..., K\}$ could be censored
- In order to estimate the parameters for the model at a certain value t=k :
- Create a matrix $X \in R^{N \times T}$ f covariates in T where t<=k (each row x1,..xT)
- Create a vector $Y \in R^{N}$ containing 1 where t=k and t is not censored, 0 otherwise
- Solve $\beta(k) = (X'X)^{-1}X'Y$

Aalen model: Matrix-factorization based regression

- Exactly the same model that was used to estimate the revenue at each level k

$$L(\hat{M}^{(k)}, \hat{N}^{(k)}) = \sum_{a \in \mathcal{D}_T^{(k)}} \gamma_1^{(T-t_a)} (C_a^{(k)} - (\hat{M}_{u_a}^{(k)})' \hat{N}_{p_a}^{(k)})^2$$
$$+ \sum_{u \in \mathbb{U}} \| \hat{M}_u{}^{(k)} - M_0^{(k)} \|_{\Omega_1}^2 + \sum_{p \in \mathbb{P}} \| \hat{N}_p{}^{(k)} - N_0^{(k)} \|_{\Gamma_1}^2$$

- It is also updated on-line after every auction

Estimating the Expected Revenue

- Gener

$$E[Rev(f)] = \int_{f}^{\infty} x d\Phi_2(x) + fP(B2 < f < B1)$$

$$= \int_{f}^{\infty} x d\Phi_2(x) + f(1 - P(B2 > f) - P(B1 < f))$$

$$= \int_{f}^{\infty} x d\Phi_2(x) + f(-P(B2 > f) + P(B1 > f))$$

Discrete case : $= \sum_{f}^{max(f)} x\phi(x) + f(-P(B2 > f) + P(B1 > f))$

Full Censorship : use the re-normalized C.D.F of B1 and B2, P(x<B1|B1<=f*)

Half Censorship : use the re-normalized C.D.F of B2, P(x<B2|B2<=f*)

Results

Real data on auctions observed during 7 consecutive days.

Protocol : 3 days data for train/development and 4 days data for testing

Metric : average revenue on the test set

The maximum possible average revenue is : 8.53

Restricting the floors to the set of discrete values : 7.87

Fixing always the floor to 0, the average revenue is : 2.59

Results

Setting 1 : Training is uncensored. Last parameters of the training phase are used to initialize the model on test phase
Setting 2 : Training is censored. Test phase parameters initialized as 0
Setting 3 : Training is uncensored. Test phase parameters initialized as 0

Method 1: Matrix Factorization + Aalen model
Method 2: Full censorship then B1=B2=0, Half censorship then B2=floor
Method 3: Matrix Factorization. In case of full censorship, do not update parameters
Method 4: Variant of Mathod 3 where floor is set by lin-UCB

	S1	S2	S3
NO_RES	2.5978	N/A	N/A
PL_RES	3.6222	N/A	N/A
M1	4.0663	3.9955	3.9957
M2	3.9012	3.7948	3.7936
M3	3.8954	3.7369	3.7468
M4	3.9306	3.821	3.8209

What is new w.r.t the state of the art?

- Matrix factorization approach
- Dealing with censored variables
- Adaptive system