

Lec3: New Directions and Opportunities in Auction Design

Guest Lectures at ZJU Computer Science
(Summer 2024)

Instructor: Haifeng Xu



Outline

- “Learning” Optimal Auction (using one sample)
- Auctioning LLM-Generated Contents
- Challenges of Selling Data Sets

Recap: Optimal Auction for Single Item

Theorem. For single-item allocation with **regular** value distribution $v_i \sim f_i$ independently, the following auction is BIC and optimal:

1. Solicit buyer values v_1, \dots, v_n
2. Transform v_i to “virtual value” $\phi_i(v_i)$ where $\phi_i(v_i) = v_i - \frac{1-F_i(v_i)}{f_i(v_i)}$
3. If $\phi_i(v_i) < 0$ for all i , keep the item and no payments
4. Otherwise, allocate item to $i^* = \arg \max_{i \in [n]} \phi_i(v_i)$ and charge him the minimum bid needed to win, i.e., $\phi_i^{-1}(\max(\max_{j \neq i^*} \phi_j(v_j), 0))$.

- Recall, “regular” means $\phi_i(v_i)$ is monotone non-decreasing
- Will always assume distributions are regular and “nice” henceforth

An important special case: $v_i \sim F$ i.i.d.

- Second-price auction with reserve $\phi^{-1}(0)$ is optimal
 1. Solicit buyer values v_1, \dots, v_n
 2. If $v_i < \phi^{-1}(0)$ for all i , keep the item and no payments
 3. Otherwise, allocate to $i^* = \arg \max_{i \in [n]} v_i$ and charge him the minimum bid needed to win, i.e., $\max(\max_{j \neq i^*} v_j, \phi^{-1}(0))$

Intuitions about why *second-price auction with reserve* is good

- Incentive compatibility requires payment to not depend on bidder's own bid → **second highest bid is pretty much the best choice**
- Use the reserve to balance between “charging a higher price” and “disposing the item”

A Key Drawback

Optimal reserve price depends on our knowledge of distribution F

How to run an auction without knowing F ?

- This is precisely the machine learning perspective
 - ML assumes data drawn from distributions
 - The precise distribution is unknown; instead samples are given

Next: will describe an elegant solution to the above I.I.D. bidder case

A Natural First Attempt

- Since “truthful” bids $\{v_i\}_{i=1}^n$ are all drawn i.i.d. from f , these n bids can be used to estimate f
- This results in the following “empirical Myerson” auction

Empirical Myerson Auction

1. Solicit buyer values v_1, \dots, v_n
2. Use v_1, \dots, v_n to estimate an empirical distribution \bar{f}
3. Run second-price auction with reserve $\bar{\phi}^{-1}(0)$ where $\bar{\phi}$ is calculated using \bar{f} instead

Q: does this mechanism work?

No, may fail in multiple ways

Issues of Empirical Myerson

Empirical Myerson Auction

1. Solicit buyer values v_1, \dots, v_n **problematic**
2. Use v_1, \dots, v_n to estimate an empirical distribution \bar{f}
3. Run second-price auction with reserve $\bar{\phi}^{-1}(0)$ where $\bar{\phi}$ is calculated using \bar{f} instead

- Not incentive compatible – reserve depends on bidder's report
 - This is a crucial difference from standard machine learning tasks where samples are assumed to be correctly given
- Even bidders report true values, \bar{f} may not be regular
- Even \bar{f} is regular, $\bar{\phi}^{-1}(0)$ may not be close to $\phi^{-1}(0)$
 - Depend on how large is n , and shape of f

The Correct Approach

- Want to use second-price auction with an **estimated reserve**
- Lesson from previous example – if a bidder's bid is used to estimate the reserve, we cannot use this reserve for him
- Main idea: pick a “**reserve buyer**” → use his bid to estimate the reserve, but never sell to this buyer
 - I.e., we give up any revenue from the reserve buyer

Q: why only pick one reserve buyer, not two or more?

We have to give up revenue from reserve buyers, better not too many

Q: which buyer to choose as the reserve buyer?

A-priori, they are the same → pick one uniformly at random

The Correct Approach

- Want to use second-price auction with an **estimated reserve**
- Lesson from previous example – if a bidder's bid is used to estimate the reserve, we cannot use this reserve for him
- Main idea: pick a “**reserve buyer**” → use his bid to estimate the reserve, but never sell to this buyer
 - I.e., we give up any revenue from the reserve buyer

Q: how to use a single buyer's value to estimate reserve?

Not much we can do . . . just use his value as reserve

The Mechanism

Second-Price auction with Random Reserve (SP-RR)

1. Solicit buyer values v_1, \dots, v_n
2. Pick $j \in [n]$ uniformly at random as the reserve buyer
3. Run second-price auction with reserve v_j but only among bidders in $[n] \setminus \{j\}$.

Claim. SP-RR is dominant-strategy incentive compatible.

For any bidder i

- If i is picked as reserve, his bid does not matter to him, so truthful bidding is an optimal strategy
- If i is not picked, he faces a second-price auction with reserve. Again, truthful bidding is optimal

The Mechanism

Theorem. Suppose F is regular. In expectation, SP-RR achieves at least $\frac{1}{2} \cdot \frac{n-1}{n}$ fraction of the optimal expected revenue.

Remarks

- $\frac{1}{2} \cdot \frac{n-1}{n}$ is a **worst-case** guarantee
- It is possible to have a good auction even **without knowing F**
 - Though we still assumed $v_i \sim F$ i.i.d.
- The proof uses very elegant geometric argument

Outline

- “Learning” Optimal Auction (using one sample)
- Auctioning LLM-Generated Contents
- Challenges of Selling Data Sets

How to Sell LLM-Generated Texts?



Haifeng

Any tips for a fun vacation in Hawaii?



Bard

.....

 This is the thing you are looking to sell

OK, but who would be interested in these texts?

One key application: digital advertisers

How to Sell LLM-Generated Texts?

 **Haifeng**

Any tips for a fun vacation in Hawaii?

 **Bard (Airline Preferred)**

Fly to Hawaii with **Alpha Airlines** and experience the beauty of the Aloha State. We offer affordable flights to all the major islands, so you can start your Hawaiian vacation sooner..

 **Bard (Hotel Preferred)**

Experience the magic of Hawaii at **Beta Resort**, where stunning views, luxurious accommodations, and endless activities await...

Mechanism Design

Goal: generate text that accounts for different agents' preferences
AND strengths of their preferences

How to Sell LLM-Generated Texts?

 **Haifeng**

Any tips for a fun vacation in Hawaii?

 **Bard (Airline Preferred)**

Fly to Hawaii with **Alpha Airlines** and experience the beauty of the Aloha State. We offer affordable flights to all the major islands, so you can start your Hawaiian vacation sooner..

 **Bard (Hotel Preferred)**

Experience the magic of Hawaii at **Beta Resort**, where stunning views, luxurious accommodations, and endless activities await...

Hopefully, via mechanism design

 **Bard (with aggregated preferences)**

Fly to paradise with **Alpha Airlines** and experience the magic of Hawaii at **Beta Resort**. Stunning views, luxurious accommodations, and endless activities await. Book your dream vacation today and create unforgettable memories.

Major New Challenges

- **Challenge 1:** functionable “allocation” rules
 - Classically, allocation → **give items** to people
 - Here, allocation → generate a **new token distribution** that needs to still function well on text generation
- **Challenge 2:** design payment that aligns with bidder preferences
 - Classically, preference → **how much** you value an item
 - Here, preference → **preferred languages??**

Hopefully, via mechanism design



Bard (**with aggregated preferences**)

Fly to paradise with **Alpha Airlines** and experience the magic of Hawaii at **Beta Resort**. Stunning views, luxurious accommodations, and endless activities await. Book your dream vacation today and create unforgettable memories.

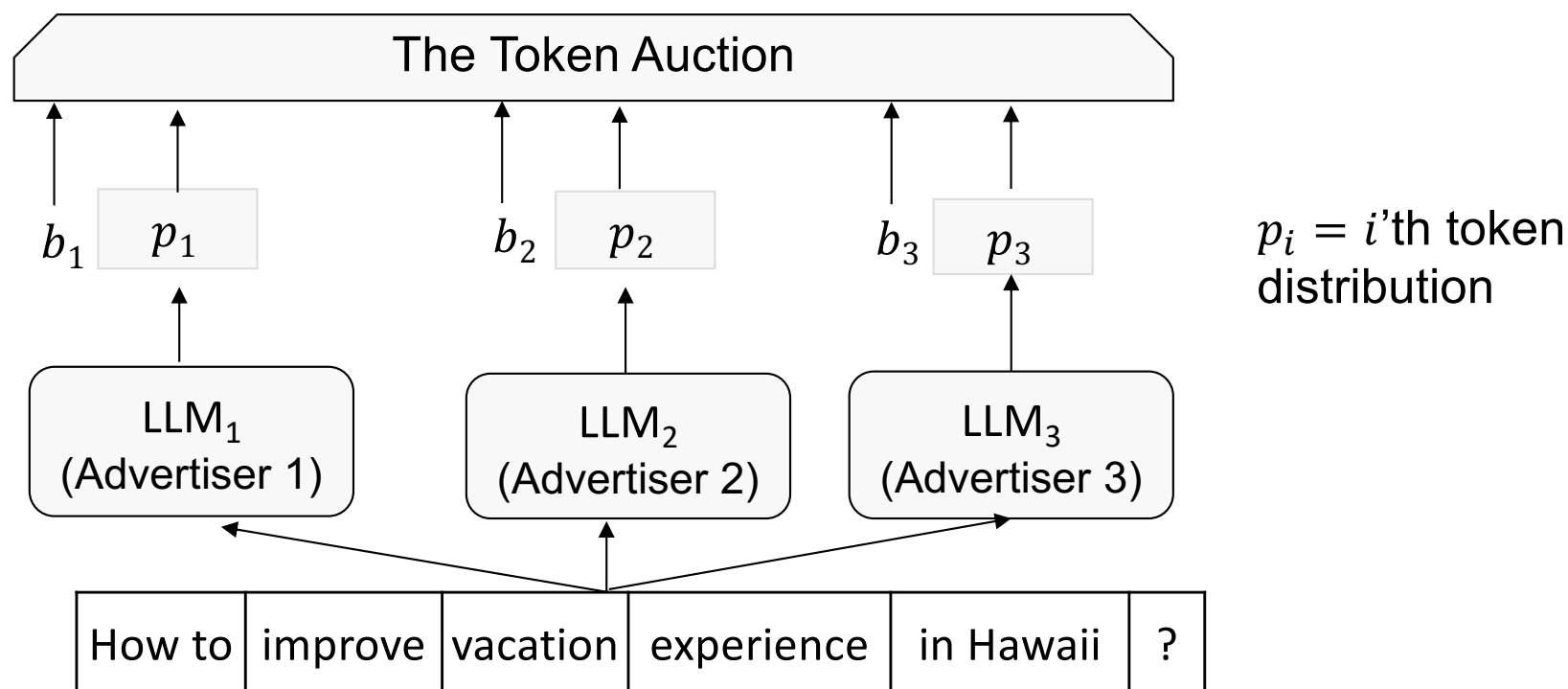
Our Solution: the Token Auction

- Advertisers are represented by LLMs
 - LLM_i encodes i 's text preferences



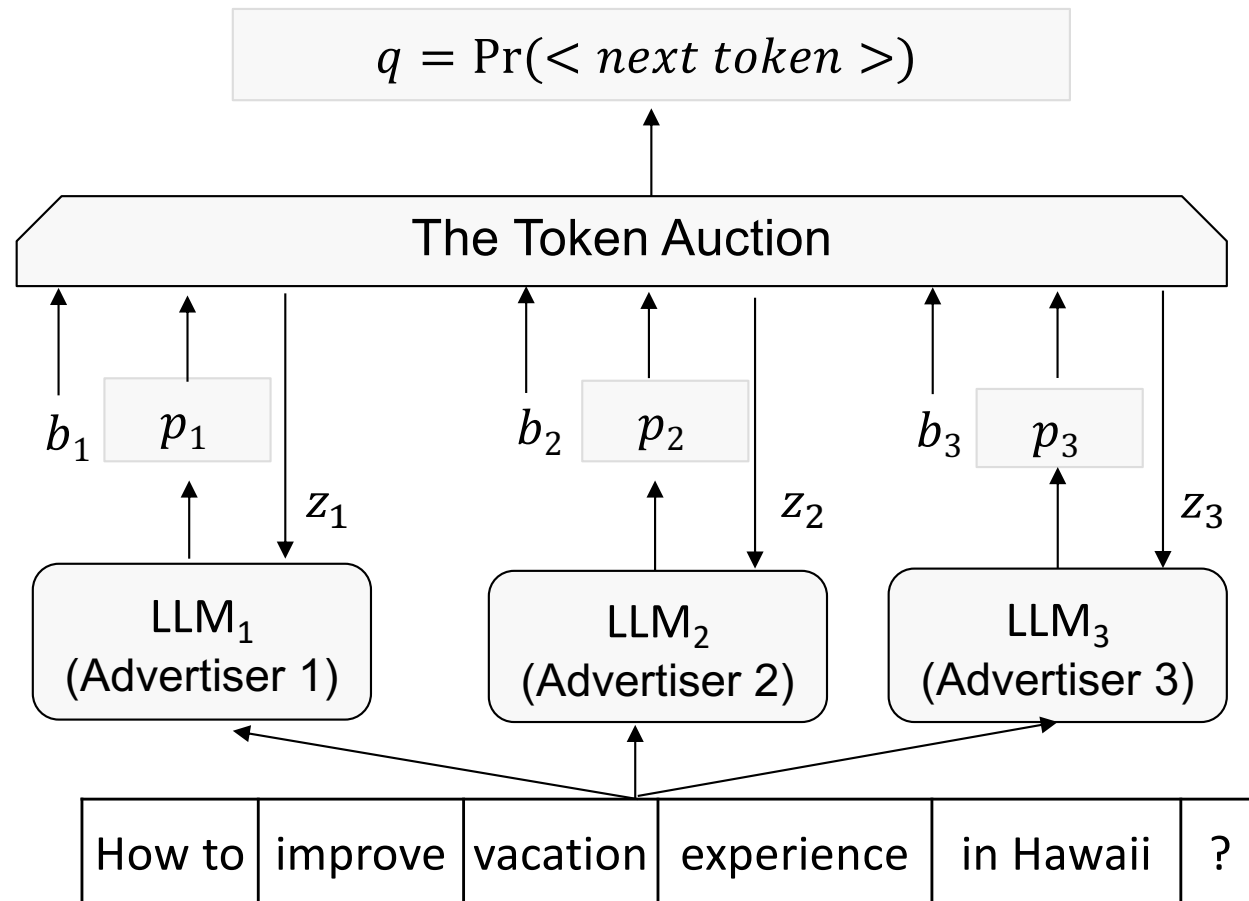
Our Solution: the Token Auction

- Auction runs only when a commercial opportunity is detected
- Each advertiser submits a bid to express their **intensity of interest**

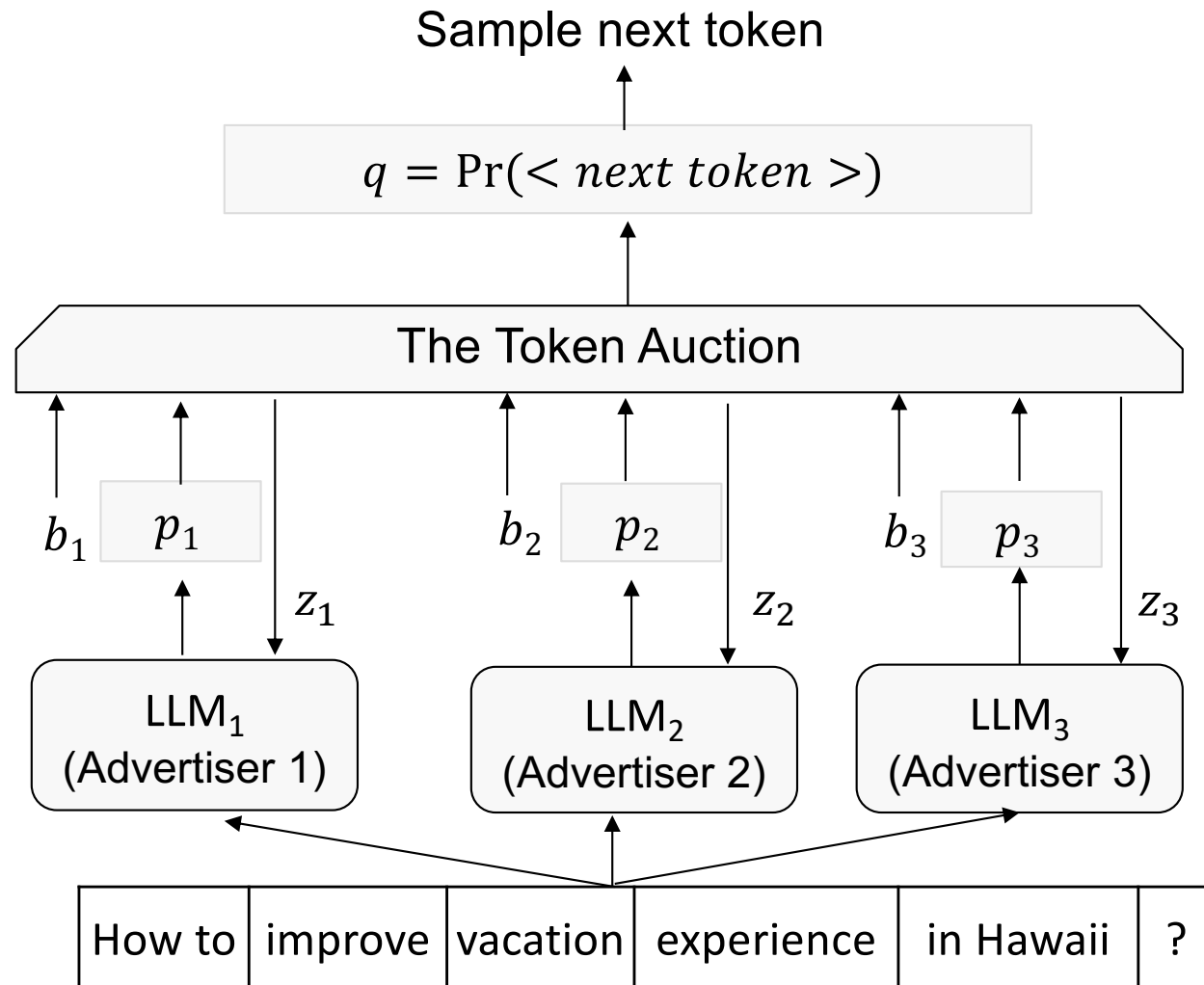


Our Solution: the Token Auction

- Auction determines aggregated distribution $q(\mathbf{b}; \mathbf{p})$ and payment $z_i(\mathbf{b}; \mathbf{p})$



Our Solution: the Token Auction



In addition to this auction format, more subtle is the concrete choice of $q, z_i \dots$

- How to describe advertisers' preferences over token distributions, and how to elicit their preferences?
- How to design aggregation function $q(\mathbf{b}; \mathbf{p})$ to “blend” their preferences?
- How to design payments $z_i(\mathbf{b}; \mathbf{p})$ to ensure strategy proofness during bidding?

See our recent paper for concrete design choices [WWW 2024]

[Mechanism Design for Large Language Models](#)



Demonstration on Google's Bard

Airline bid: **1**

Hotel bid: **0**



Haifeng

Any tips for a fun vacation in Hawaii?



Bard

Alpha Airlines: Your ticket to paradise.

Demonstration on Google's Bard

Airline bid: **0.75**

Hotel bid: **0.25**



Haifeng

Any tips for a fun vacation in Hawaii?



Bard

Alpha Airlines: Fly to Hawaii and experience the beauty of the islands with aloha.

Beta Resorts: Stay at our resorts and enjoy the best of Hawaii.

Demonstration on Google's Bard

Airline bid: **0.25**

Hotel bid: **0.75**



Haifeng

Any tips for a fun vacation in Hawaii?



Bard

Experience the magic of Hawaii at the Beta Resort, where the sun shines brighter and the waves crash louder — book your stay today with our exclusive 20% off discount!

Issues of hallucination is here!

Demonstration on Google's Bard

Airline bid: 0

Hotel bid: 1



Haifeng

Any tips for a fun vacation in Hawaii?



Bard

Hawaii's Beta Resort: a paradise where the sun shines brighter, the waves sing sweeter, and the sand feels softer.

While this study is motivated by digital advertising, it is applicable to other preference aggregation to meet social choices

Outline

- “Learning” Optimal Auction (using one sample)
- Auctioning LLM-Generated Contents
- Challenges of Selling Data Sets

We Already Know How to Sell Items



How Is Selling Data Different?



Data can be curated via complex and **highly non-linear** transformation

A Simple Example

- A buyer want to estimate $\beta_1 + \beta_2$ for a 2-D parameter $\beta = (\beta_1, \beta_2)$
- Seller has two data records generated from linear model $y = \beta \cdot x + noise$

$$x^a = (1, 2), y^a = 3$$

$$x^b = (2, 1), y^b = -1$$

- Can sell the raw data but can also sell processed data:

- ✓ Average $\bar{x} = \frac{x^a + x^b}{2}, \bar{y}$
- ✓ Difference $x^\perp = \frac{x^a - x^b}{2}, y^\perp$
- ✓ Even predicted β
- ✓ ...



Most complex allocation is just to use lottery

How Is Selling Data Different?



Data can be curated via complex and **highly non-linear** transformation

Should sell **data** or data-enabled **prediction**?

Complex preferences: context dependent, externalities, uncertainty

Privacy concerns further complex all above




Most complex allocation is just to use lottery

No question of whether to sell a house or its building materials

Just a value for a set of items

Typically no such constraints

How Is Selling Data Different?



How to address these challenges in an optimal way is still mostly open.

Data can be curated via complex and **highly non-linear** transformation

Most complex allocation is just to use lottery

Should sell **data** or data-enabled **prediction**?

No question of whether to sell a house or its building materials

Complex preferences: context dependent, externalities, uncertainty

Just a value for a set of items

Privacy concerns further complex all above

Typically no such constraints

Thank You

Haifeng Xu

University of Chicago

haifengxu@uchicago.edu