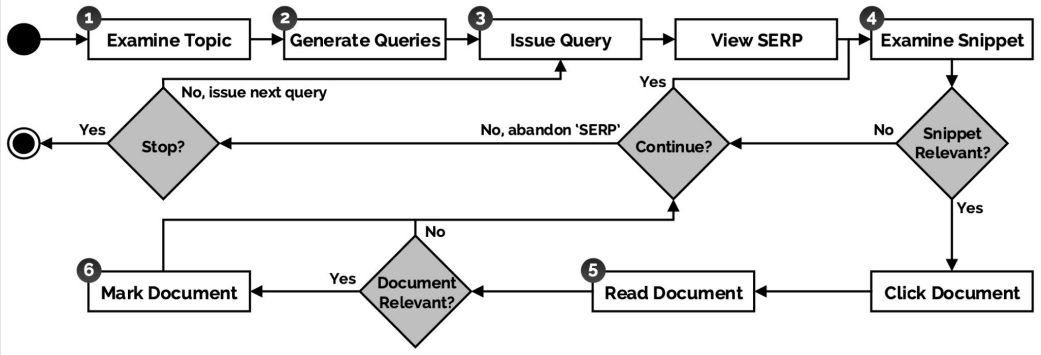
03

For Information Retrieval

Search Engine
Conversational System

From Complex Searcher Model to Search-Engine Simulation

[Maxwell, Leif 2016]



Gery components can be characterized by agent decisions

White components can be characterized by agent models

Basic setup:

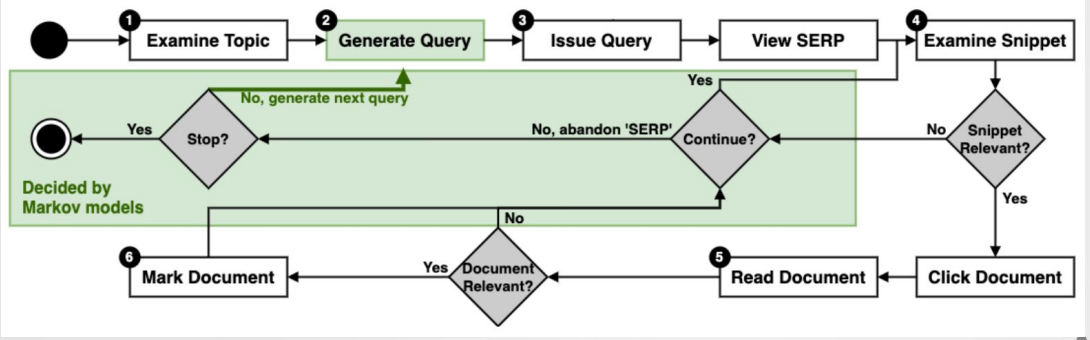
- Query agent model (feature model)
- Document agent model (feature model)
- User agent model (feature model, choice model)
- Environment (stopping criteria, search context)
- ……

Goal: 'sim-to-real' scenario for system evaluation

Extend Basic CSM Setup to Accommodate More Scenarios and Design Strategies

[Zerhoudi, 2022]

Controlled user behavior by indexing user agent model with user type



Markovian query generation by adding a transition model to query agent

Adding controlled behavior for the user agent:

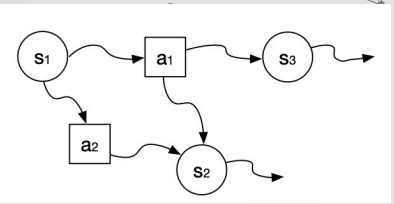
- Stylized system evaluation
- User-centric and beyond-average analysis
- Opportunity discovery for different user types

Adding memory and path dependencies to query agent:

- Fine-grained control over environment complexity
- Reflect complex agent decision and interaction
- Faithfulness vs. manipulability

Agenda-based User Simulation for Dialogue System

[Schatzmann 2007]



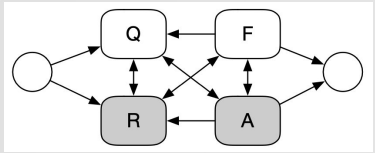
- **User agent** (goal model, agenda model, agenda act model, transition model)
- **Environment** (+ partially observed MDP)

User simulation at a semantic level	*Goal- and Agenda-Based State Representation*	*User act selection*	*State transition model*	*Agent update model*	*Goal update model*
Current state | User action | Intermediate state | System action | New state	A. User behave in a consistent and goal-driven fashion B. Agenda as a stack-like structure containing pending user acts (inform, request)	Pop items from the stack Can be made stochastic	• (Stochastic) push operations where dialogue acts are added to the agenda • The hidden user constraints and requests changes with a given machine action • MDP update		

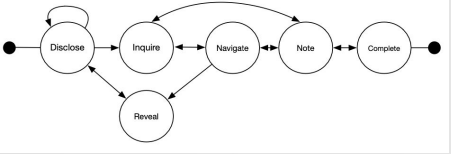
Conversational System Simulation

[Zhang, Krisztian 2020]

- **User agent** (goal model, agenda model, agenda act model, transition model)
- **+ Interaction model, Preference model, Natural language agents**
- **Environment** (+ partially observed MDP)
- **+ conversational agent**



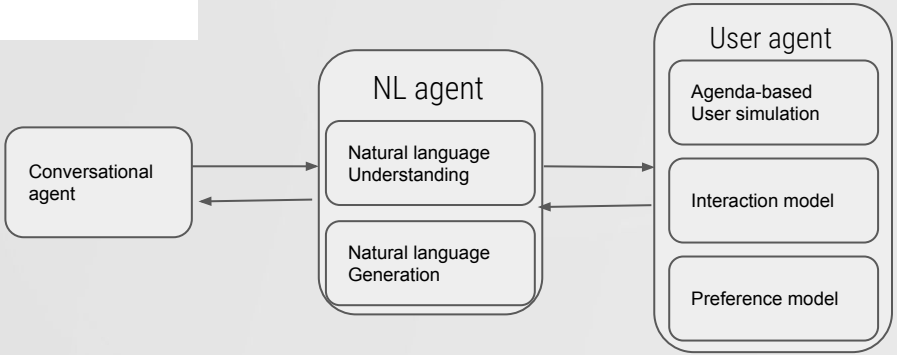
QRFA model
(Query, Request, Feedback, Accept)



CIR6 model
(Conversational item rec model)

Interaction model

Preference model
- Single Item Preference model
- Personal knowledge graph



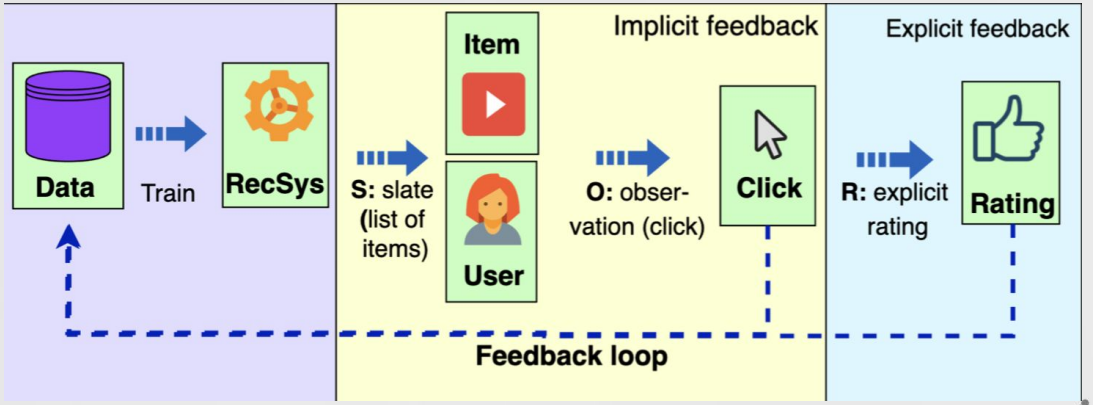
Template-based NL models
- Natural language understanding
- Natural language generation

04

For Recommender System

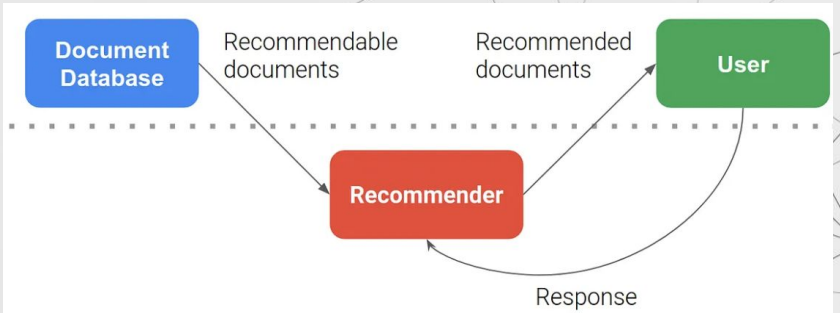
Personalized Recommendation

Data Generation Process in RS



[Stavinova et al. 2022]

- **Users and Items Models**: Models for generating synthetic profiles of users and items.
- **Recommender Systems**: Models for recommending items to users.
- **User Response Models**: Models for providing feedback.



RecSim [Ie et al. 2019]

Simulator Design

Recommender Systems

- The construction of RecSys depends on scenarios and assumptions.
- Could be RL agent or pre-defined models.

Goal

- Scenarios: specific task and scenarios that could take place in the interaction between users and items
- Assumptions: a set of assumptions about the mechanism to satisfy the scenarios.

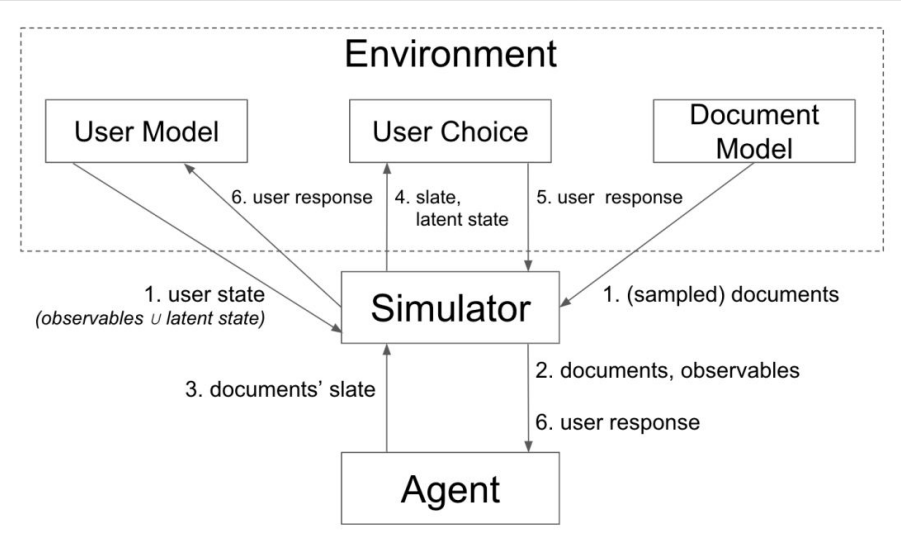
Users/Items Profile

- Realism: The profile can be generated with or without real data.
- Uncertainty: The profile may include noise and uncertainty.
- Dynamics: The profile can be dynamic or static.

User Response Models

- Related on the scenarios and assumptions..
- Based on user/item features, history, context, etc.
- Generate implicit/explicit feedback, time, etc.

RL-based Simulator Pipeline

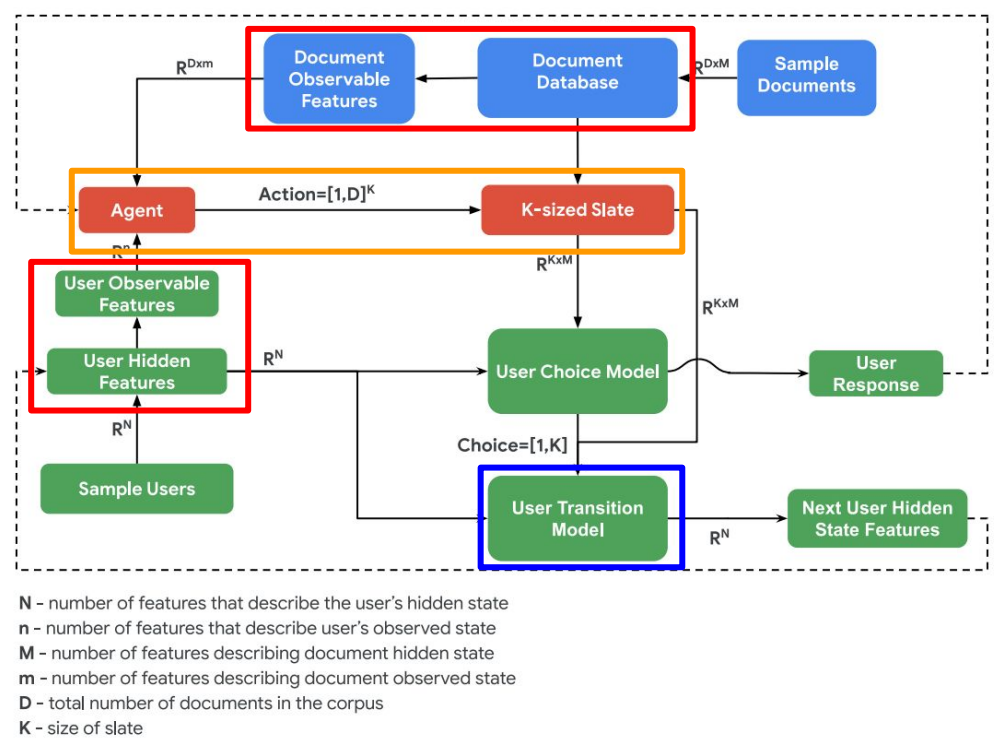


RecSim [Ie et al. 2019]

- The **environment** consists of a user model, a document (item) model and a user-choice model.
- The **agent** serves as a recommender system.
- The **action** is defined as recommending item(s).
- The **reward** will generally be a function of a user's responses

Case Study: RecSim

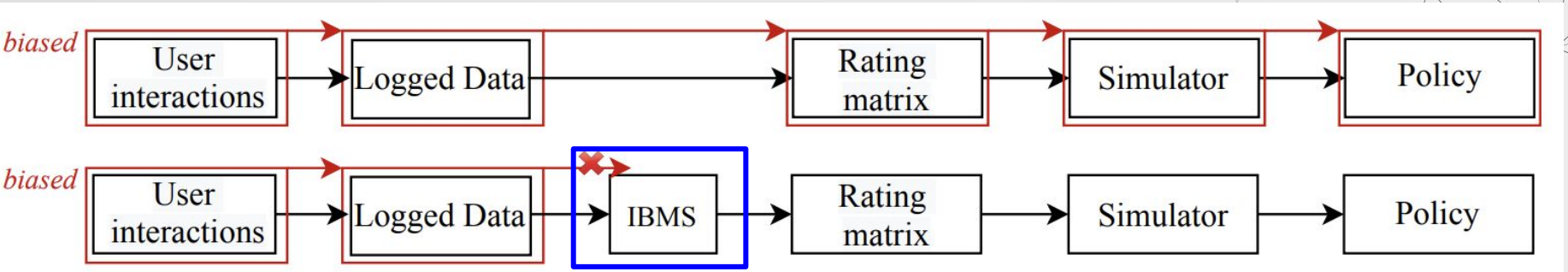
[Ie et al. 2019]



- **Scenario**: providing environments that facilitate the development of new RL algorithm for recommender applications (Collaborative Interactive Recommenders). 'Sim-to-real' is not the priority concern.
- **User/Item profile**: sampled from a prior (based on real data or not) over user/item features, including both latent and observable features
- **Dynamics**: User profile will be updating along with interactions by User Transition Model
- **RecSys**: act as an agent to recommend slates of documents (items) based on observed features.
- **User Response Model**: generate user response depending on observable item features and all user features (latent and observable)

Manipulate the Simulator Generation

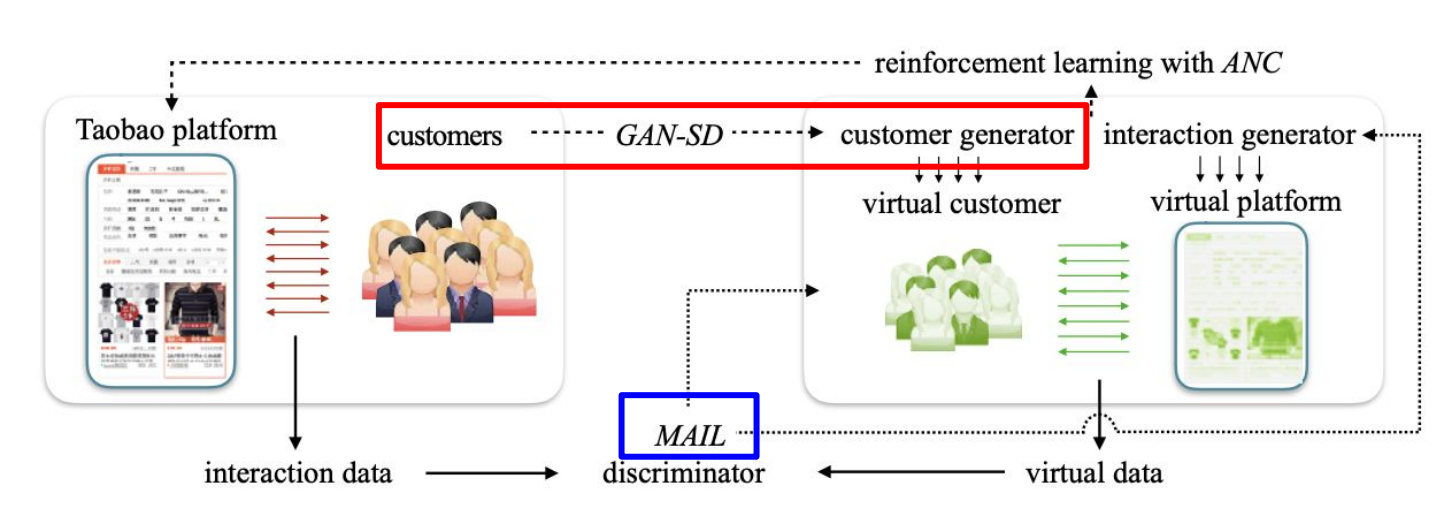
[Huang et al. 2020]



- **Scenario**: interactive recommender.
- **Concerns**: Using real data to build simulator may suffer from bias of real data.
- **User Response Model:** Predicted rating matrix
- **Intermediate Bias Mitigation Step (IBMS):** mitigating the effect of bias before the prediction model is learned by IPS

Case Study: Virtual-Taobao

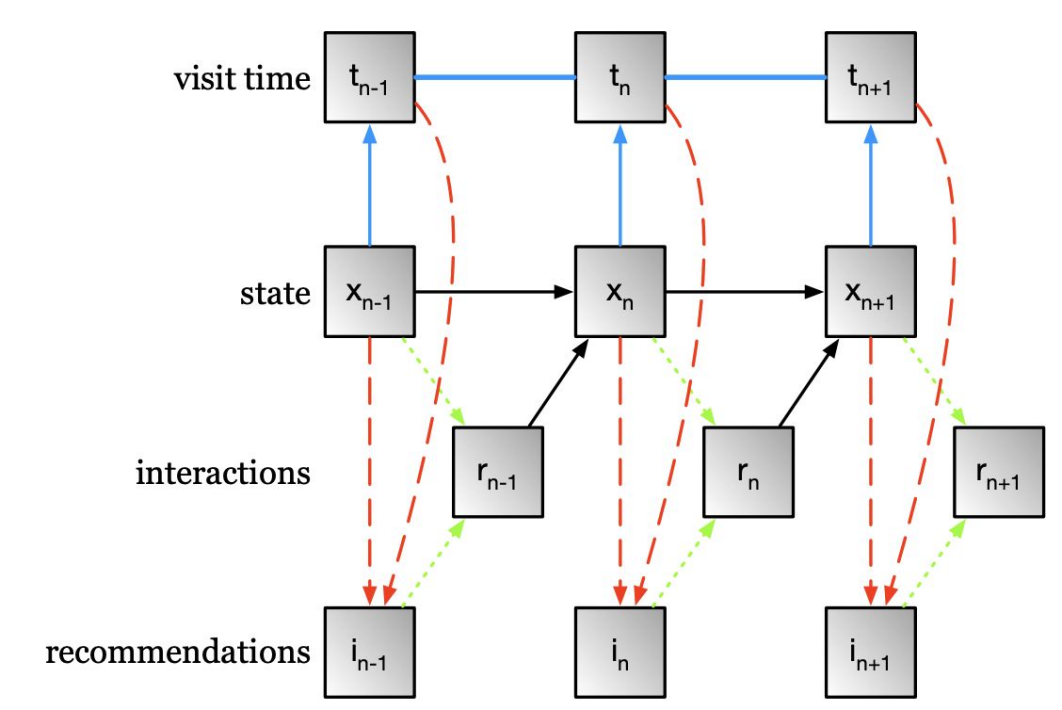
[Shi et al. 2018]



- **Scenario**: Real-world Online Retail Environment (Taobao)
- **User profile**: focus on realism, GAN-SD (GAN Simulation Distribution) to generate consumers similar to real data.
- **Interactions**: Generated by MAIL (multi-agent adversarial imitation learning), training the **customer policy** as well as the **engine policy**.

Case Study: Accordion

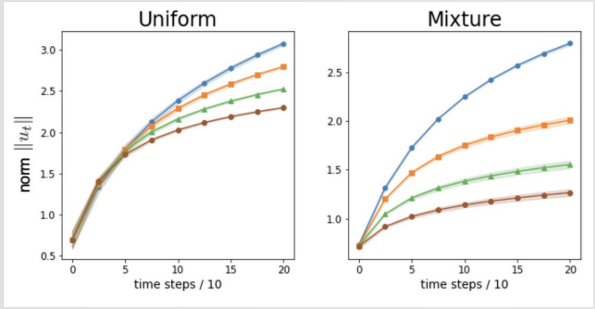
[McInerney et al. 2021]



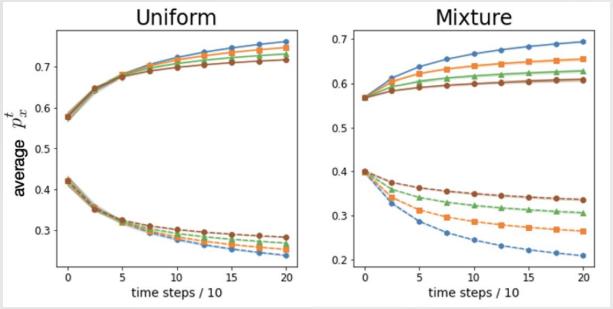
- **Scenario**: simulating Long-term interactive systems. Time-aware recommendation scenario
- **RecSys**: Recommendation models that considering visit time and user state for making recommendations.
- **User Response Models**: Consisted of visit model and selection model.
- **Visit Model**: A Poisson process based method for simulating user visit, time of the visits and the number of interactions in each visit.
- **Selection Model**: simulating outcome of the interaction (e.g., click, purchase, stream)

Simulation for Observation

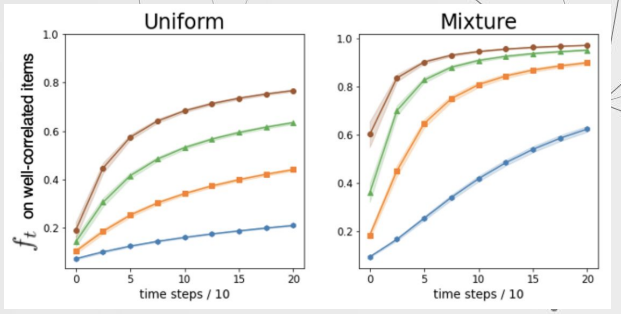
[Kalimeris et al. 2021]



user preference



probability of likable vs non-likable item



Probability mass on items correlating well with the initial user preference

- **Scenario**: study the preference amplification caused by MF models
- **User profile**: sampled from uniform distribution. Updated after every interactions.
- **Item profile**: sampled from pre-defined distributions (uniform or mixture of two uniform)
- **RecSys**: MF-based models.
- **Conclusion**: preference amplification, echo chambers, filter bubbles.

Manipulate User Response Models for Observation

[Yao et al. 2021]

	Name	Formula
Choice	Lazy	$\mathcal{M}_s^{lazy}(v) = \mathbb{1}[\text{Rank}(v) = 1]$
	Uniform	$\mathcal{M}_s^{uniform}(v) = \frac{1}{k}$
	Ranked	$\mathcal{M}_s^{ranked}(v) \propto \frac{1}{\log(1+\text{Rank}(v))}$
	α-preference	$\mathcal{M}_s^{\alpha-prefer}(v) \propto e^{\alpha\rho(v)}$
Feedback	Positive	$\mathcal{M}_f^{positive}(v) = +1$
	β-preference	$\mathcal{M}_f^{\beta-prefer}(v) = \begin{cases} \beta & \rho(v) \geq \rho_0 \\ -\beta & \text{else} \end{cases}$

- **Scenario**: measuring the impact of a recommender system (popularity bias) under different types of user behavior.
- **User profile**: trajectory obtained from the real data
- **Item profile**: item features obtained from the real data (popularity denoted as \rho(v))
- **User Response Models**: Consisted of choice model (Implicit) and feedback model (binary explicit).
- **RecSys**: Pre-trained MF and RNN.

05

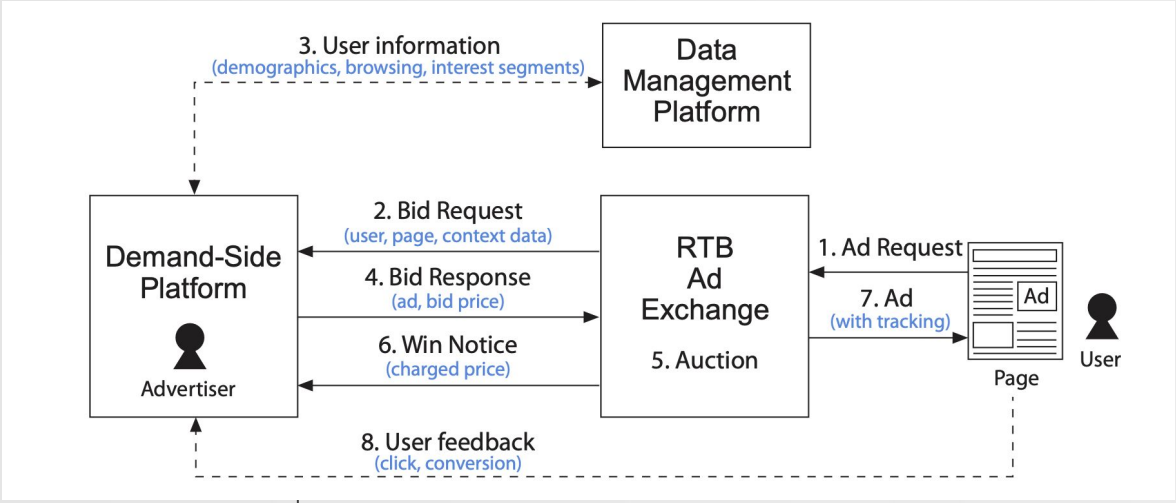
For Marketing and Advertising

Bidding, Pricing, Ads Allocation

How RTB(Real-Time Bidding) Works

[Weinan, 2014]

- **Advertisers** create advertisement campaigns and place bids that describe how much they are willing to pay to see their ads displayed or clicked.
- **Publishers** provide web services. They provide the infrastructure to collect advertiser bids and display selected ads and expect to receive payments from the advertisers.
- **Users** reveal information about their current interests. They are offered web pages that contain a selection of ads. Users may view/click/place an order on an advertisement.



Auction Mechanism

Myerson [1981] Auction is generally regarded as a fair and transparent way for advertisers and publishers to agree with a price quickly, whilst enabling the best possible sales outcome.

- ***Publishers*** have access to partial information about the market demand from historic transactions. However, they do not have knowledge about how much an individual ad impression is worth on the market.
- ***Advertisers*** may have different (private) valuations of a given ad impression.

Second-price Auction.

- Truthful bidding is a dominant strategy in second-price auctions under several assumptions:
  - Bidder knows their expected valuation given a context
  - Placed bids do not influence the value of the good
  - Competitors all have access to the same information
  - Repeated rounds of auctions are statistically independent

First-price Auction.

- Bidders should optimally shade their bids to balance the trade-off between paying lower prices and decreasing their chances of winning.

Causation Issues in Computational Advertising

With the development of new ads marketplace algorithms there are always 'what if' questions with any policy, parameter or model change in the system that yields a different ad allocation.

Controlled Experiments (Kohava [2008])

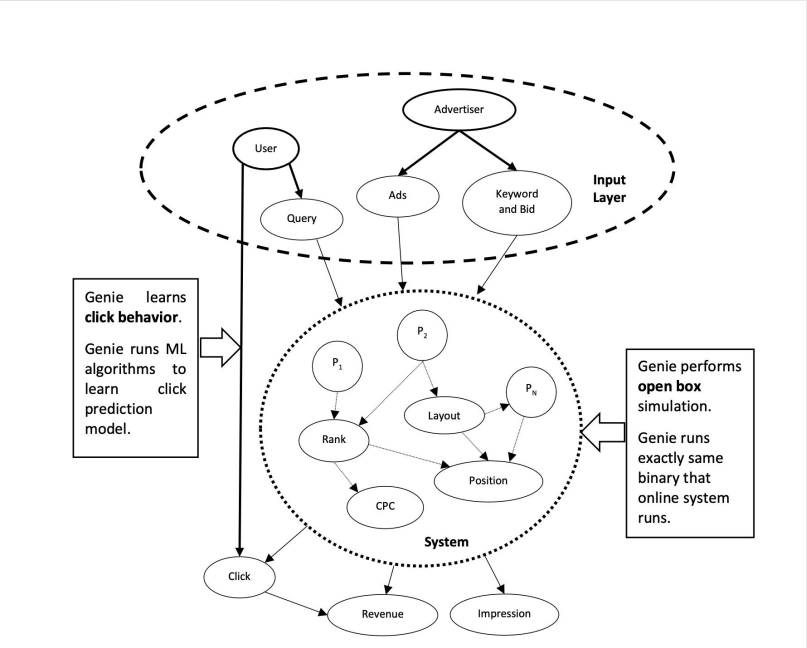
- **Expensive** because they demand a complete implementation of the proposed modifications.
- **Slow** because each experiment typically demands a couple months.
- Splitting advertisers into treatment and control groups demands special attention because each auction involves multiple advertisers. **Simultaneously controlling for both users and advertisers is probably impossible**.

Statistical Methods (Simpson [1951])

- **Cheaper and faster** statistical methods are needed to drive essential aspects of the development of RTB engine. However, interpreting cheap and fast data can be very deceiving.
- **Confounding Data**: Assessing the consequence of an intervention is generally challenging because of difficulty to determine whether the observed effect is a consequence of the intervention or has uncontrolled causes.

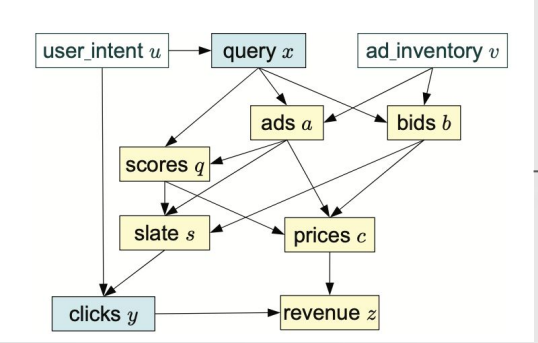
Modeling Causal Systems

Bayir [2019] proposes counterfactual policy estimation framework called Genie to optimize Sponsored Search Marketplace. Genie employs an open box simulation engine with click calibration model to compute the KPI impact of any modification to the system.

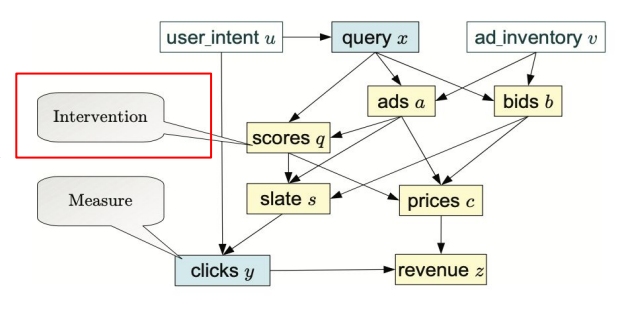


- **KPI impact** of any policy can be estimated by replaying training data with the modified policy and using user click behavior model that has tolerable noise.
- **Explore much wider parameter space** since it does not require real traffic with modification/exploration cost.
- Be leveraged to **tune completely new policies** where creating initial experiment is very costly due to cold start problem.

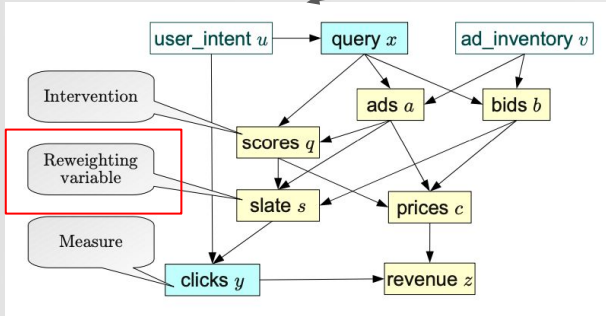
COUNTERFACTUAL REASONING AND LEARNING (Bottou [2013])



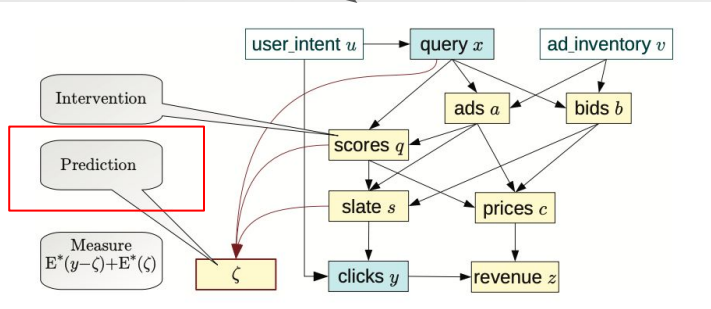
Casual Graph



Intervention



Displace reweighting point



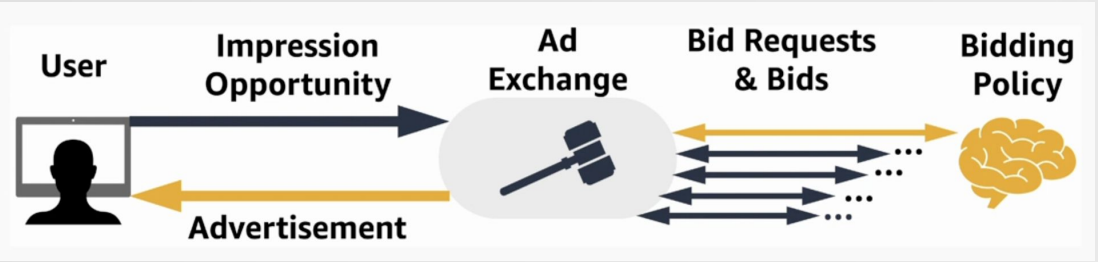
Use prediction point

Bandit learning for bidding strategies

Olivier [2022] introduce AuctionGym: a simulation environment that enables the use of bandit learning for bidding strategies in online advertising auctions.

Simulating auctions end-to-end

- an impression opportunity arises with features x~P(X)
- auction presents this opportunity to bidders
- bidders decide on an ad to show and a bid to place
- auctioneer decides on the auction winner and price
- the winning ad is shown and conversion/click/impression is observable by the winning bidder



Auction Gym

- Policy-based and doubly robust formulation of bidding problem
  - Interactive and reactive nature of the repeated auction mechanism.
- Bandit-based “learning to bid”
  - Ad allocation problem
  - Bidding problem.

Auction Gym

- **Simulating Auctions (First/Second price auctions) to decide:**
  - who wins the auction
  - how much they will be charged
- **Simulating Bidders**
  - Every bidder has a private ad catalogue, private valuation for a given ads on conversion event. The ad-specific parameters are configurable.
- **Simulating Advertising Outcomes**
  - Simulate whether an allocation decision leads to a conversion event for the advertiser.

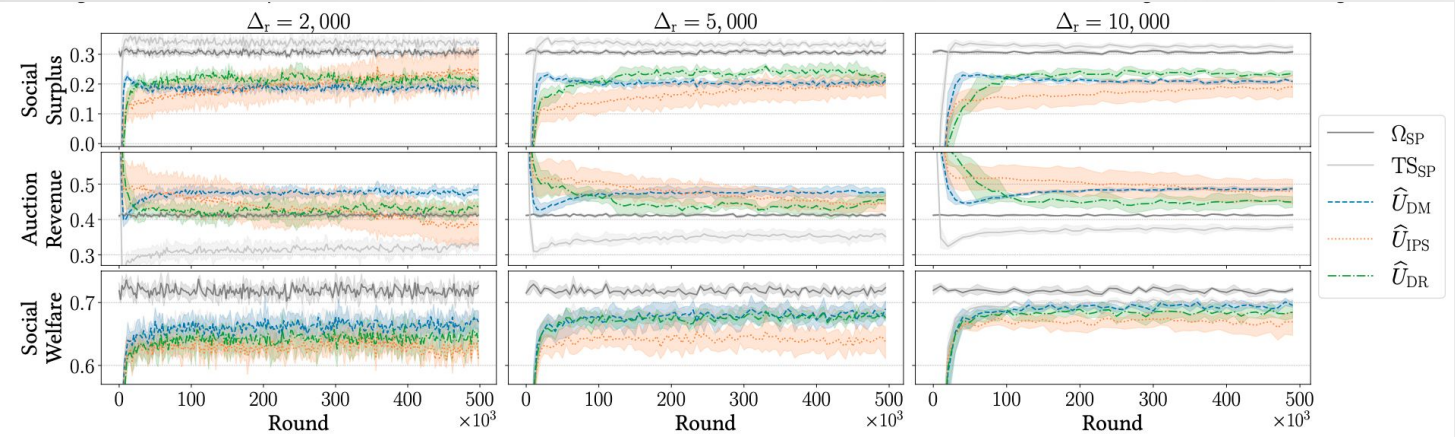
Off-policy estimation

- Choose a counterfactual (off-policy) estimator:
  - Given samples from π_0, what utility would I get from π?
- Learn the policy that maximizes this estimator: optimize π through gradient descent

- Value-based estimation (Direct Method)
  - Model winning probability function
  - high bias
- Policy-based estimation (IPS)
  - High variance
- Doubly robust estimation
  - Unbiased, lower variance

Case Study: Auction Gym

- **Model-based approach** stabilises quickly but suboptimally
  - Biased low-variance estimator.
- **Model-free importance sampling estimator** has high variance, and is able to improve upon the model-based estimator when sufficient learning steps are allowed.
  - The instability can lead to significant reductions in attainable welfare as it impacts training data collection for subsequent updates to the allocation model.
- **Doubly robust estimator** leads to improved surplus over all bidders participating in the auction
  - Lower variance than IPS.



Challenges and Future Directions

Challenges

- Online vs offline parity
  - Tuning setup has significant deviations from existing policies/models in real traffic, yield large change in feature distributions.
- Increasingly large data size and search space
  - Data size grows aggressively including traffic volume, ads data and contextual data. Increasing complexity of the problem space need calibration on critical steps.

Future Directions

- Full reinforcement learning
  - Full reinforcement learning instantiations of the bidding problem, where current actions influence future states and a notion of planning can further improve bidder surplus
- Extend the simulation environment
  - Support advertiser budgets, multi-item and learnt auction mechanisms.