LinUCB in Recommendation

Radek Bartyzal

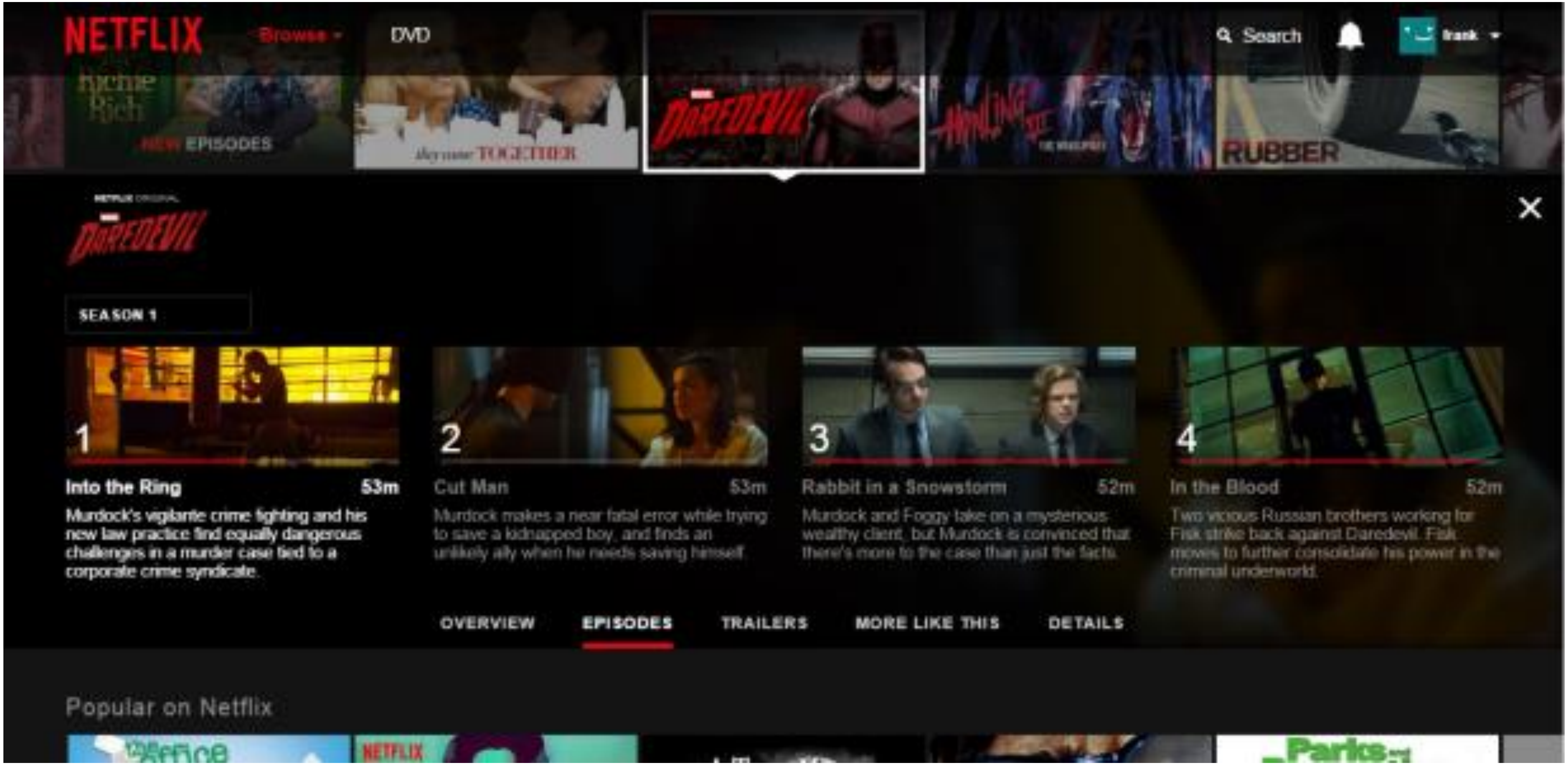
MI-GLR

25. 5. 2018

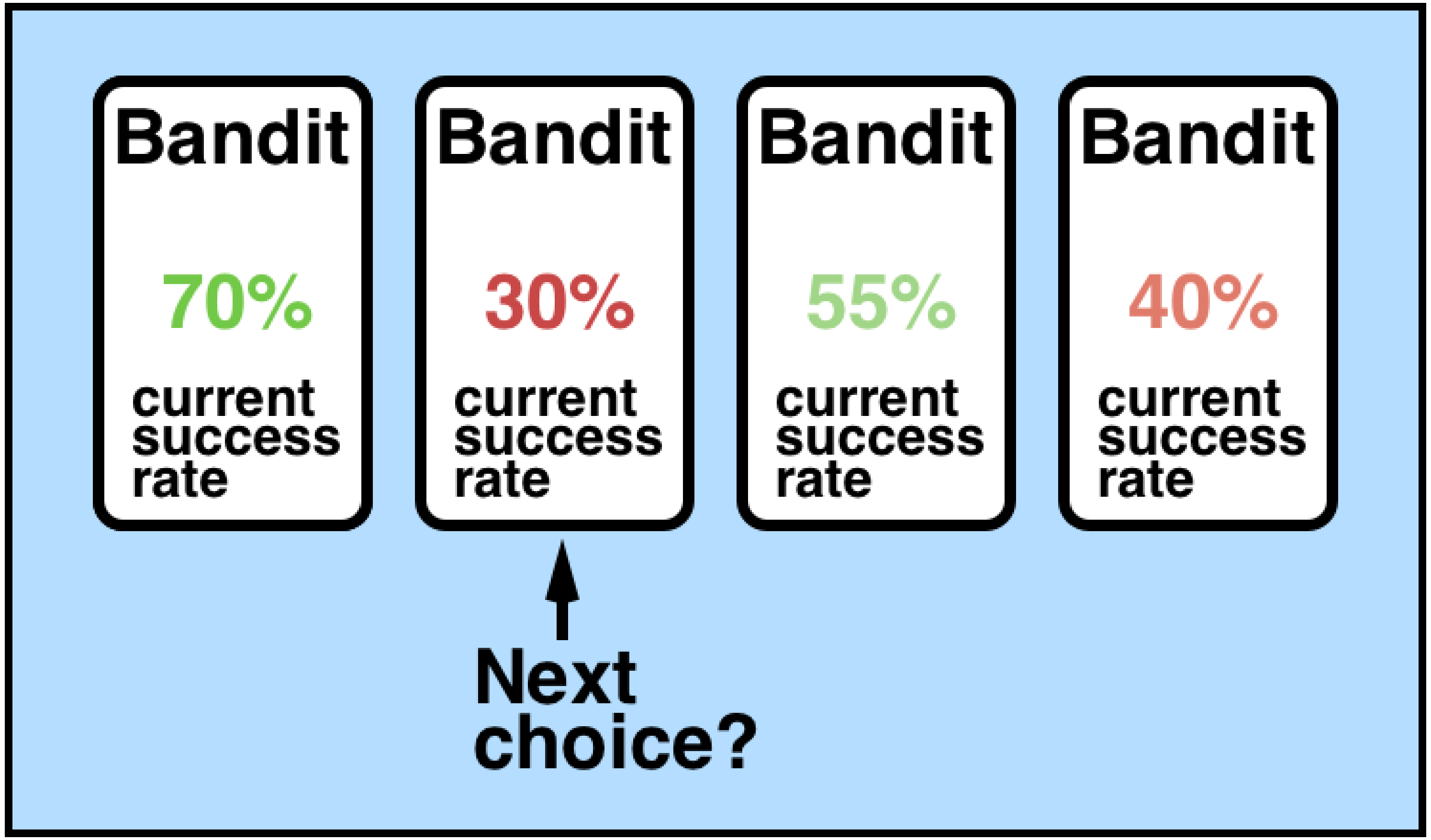
Recommender system: ratings

$users/items$	i_1	i_2	$\cdots$	$\cdots$	$\cdots$	$\cdots$	i_m
u_1	$r_{1,1}$	$\cdots$	$r_{1,j}$	$\cdots$	$r_{1,k}$	$\cdots$	$r_{1,m}$
u_2	$r_{2,1}$	$\cdots$	$r_{2,j}$	$\cdots$	$r_{2,k}$	$\cdots$	$r_{2,m}$
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
u_i	$r_{i,1}$	$\cdots$	$r_{i,j}$	$\cdots$	$r_{i,k}$	$\cdots$	$r_{i,m}$
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
u_n	$r_{n,1}$	$\cdots$	$r_{n,j}$	$\cdots$	$r_{n,k}$	$\cdots$	$r_{n,m}$

Recommender system: attributes



Bandit algorithms



Contextual Bandit algorithms

- selects articles to serve users
- based on contextual information about the users and articles
- while adapting its article-selection strategy

Upper Confidence Bound (UCB)

- balance exploration and exploitation
- in trial t, estimate:
    - mean payoff **μ_t,a** of each arm
    - corresponding confidence interval **c_t,a**
    - so **|μ_t,a − μ_a| < c_t,a** holds with high probability
- select the arm that achieves a highest upper confidence bound = **μ_t,a + c_t,a**

- with right confidence intervals UCB has logarithmic regret in the total number of trials T = optimal

LinUCB with Disjoint Linear Models

We assume the expected payoff of an arm a is linear in its d-dimensional feature $x_{t,a}$ with some unknown coefficient vector θ_a^*, namely:

$$E[r_{t,a}|x_{t,a}] = x_{t,a}^T \theta_a^*$$

- a = arm = action = item to be recommended = e.g. article
- t = trial = in this case user ID
- $r_{t,a}$ = reward of action a in trial t
- $x_{t,a}$ = features describing **both** user and the selected article a at trial t. If the features of a user are his ratings, they will change with time.

LinUCB with Disjoint Linear Models

- disjoint = parameters are not shared among different arms
- solution is reached by a simple ridge regression

LinUCB with Hybrid Linear Models

- arm-specific features as with Distinct
- features shared by all arms – new
  - user may prefer only articles about politics
  - e.g. user prefers items from genre X = feature shared by all arms

$$E[r_{t,a}|x_{t,a}] = z_{t,a}^T \beta^* + x_{t,a}^T \theta_a^*$$

- $z_{t,a}$ = features of the current user/article combination
- β^* = an unknown coefficient vector common to all arms

Dataset

- MovieLens 100k dataset
- 100 000 ratings
- 1000 users
- 1700 movies
- ratings scale 1-5 -> 1 (>=4) | -1 (<4) | 0 (unknown)

- Used subset: 100 items and 56 users
- randomly added 3 ratings to each user

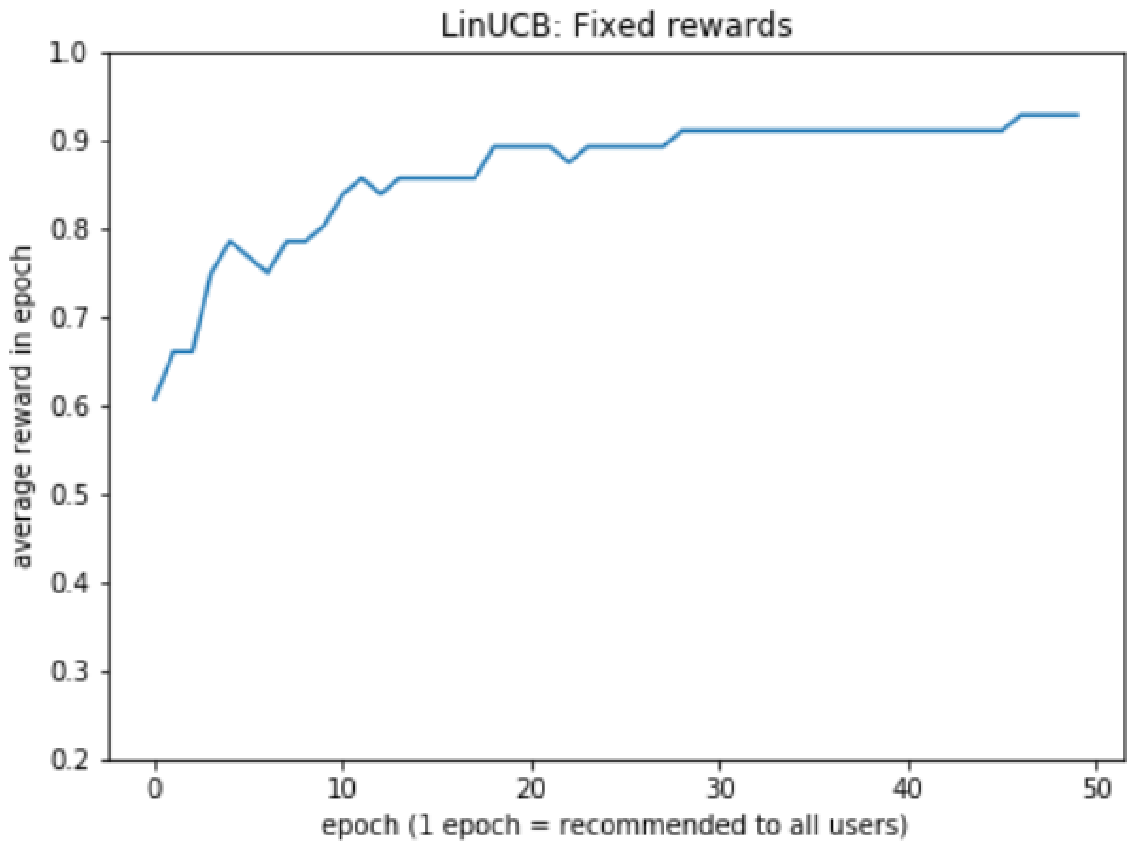
User modeling

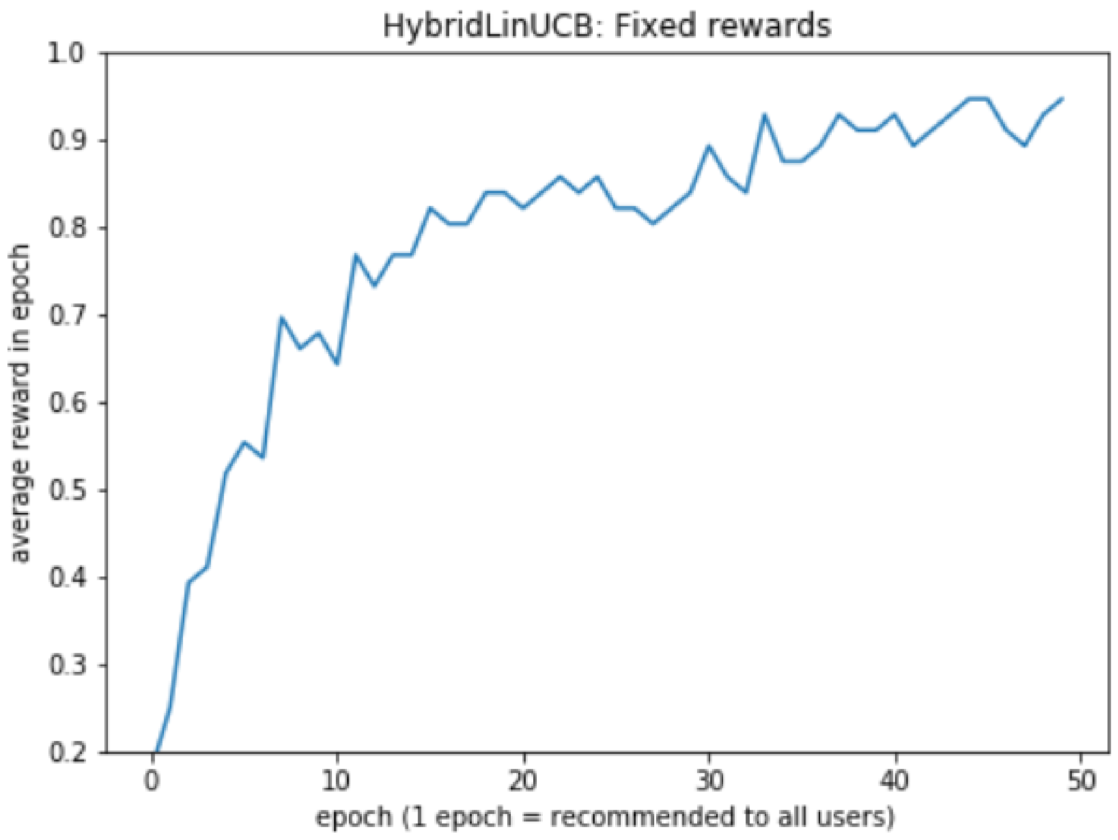
- I need to predict rating for each user/item pair = simulate reaction of real users to recommendations
- Let item a be recommended to user u:
  - If the user u has already rated the item a, the returned reward will be 1 for positive rating or 0 for negative rating.
  - Else: reward will be sampled from a Bernoulli distribution with p equal to how much the user likes the genre of item a.
  - Likability of a genre g is calculated as a ratio of positive ratings of items belonging to genre g to a number of negative ratings of items belonging to g.

Implementation

- Features in x_t,a are a concatenation of features of user t and item a at the current time
- user features = his ratings of all items = Ru
- item features = movie genres = array of 1/0 describing whether the movie belongs to a genre. There are 19 genres.
- x_t,a = user features concatenated with item features
- z_t,a = only article features = genre vector

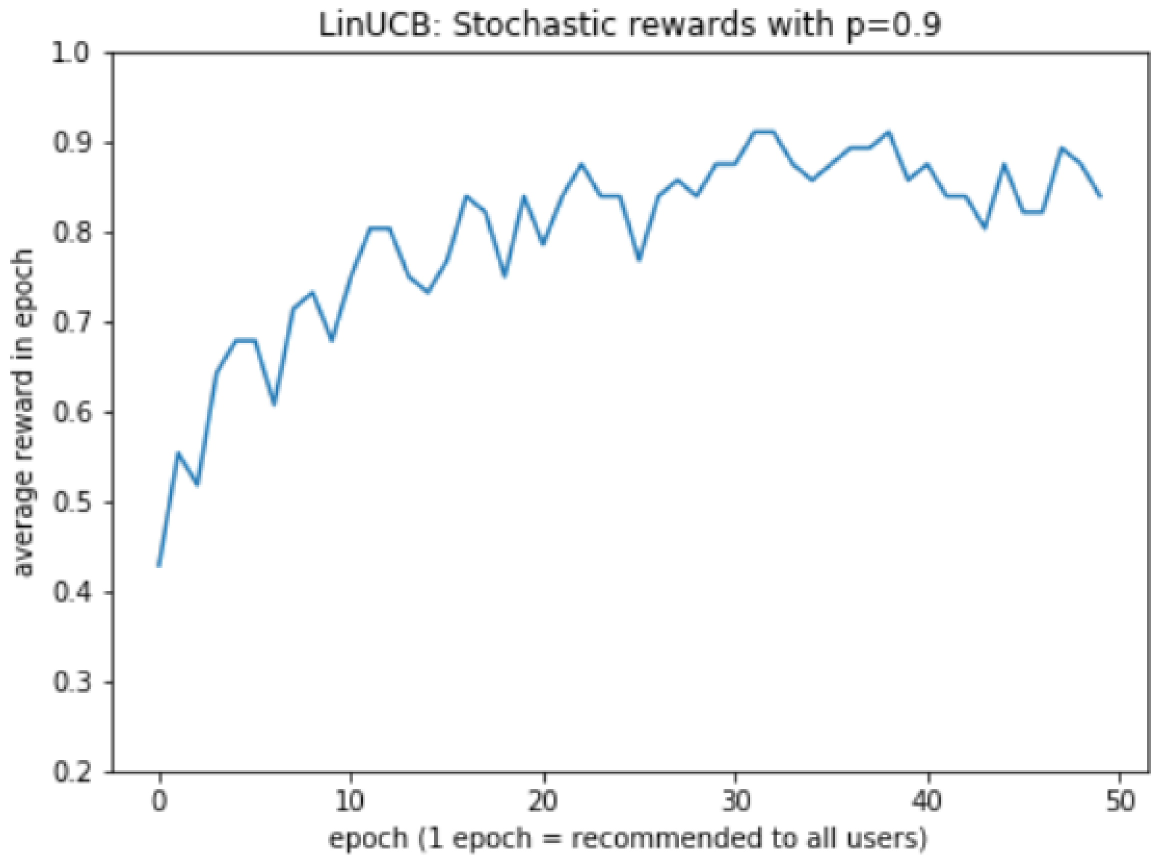
Experiment with deterministic rewards





Experiment with stochastic rewards

- Reward 1 is returned with p=0.9



Experiment with stochastic rewards