DISTRIBUTED STOCHASTIC CONTEXTUAL BANDITS FOR PROTEIN DRUG INTERACTION

Jiabin Lin[1], Karuna Anna Sajeevan[2,3], Bibek Acharya[2], Shana Moothedath[1], Ratul Chowdhury[2,3]

1. Department of Electrical and Computer Engineering 2. Department of Chemical and Biological Engineering 3. The Center for Biorenewable Chemicals
Iowa State University, Ames, IA, USA

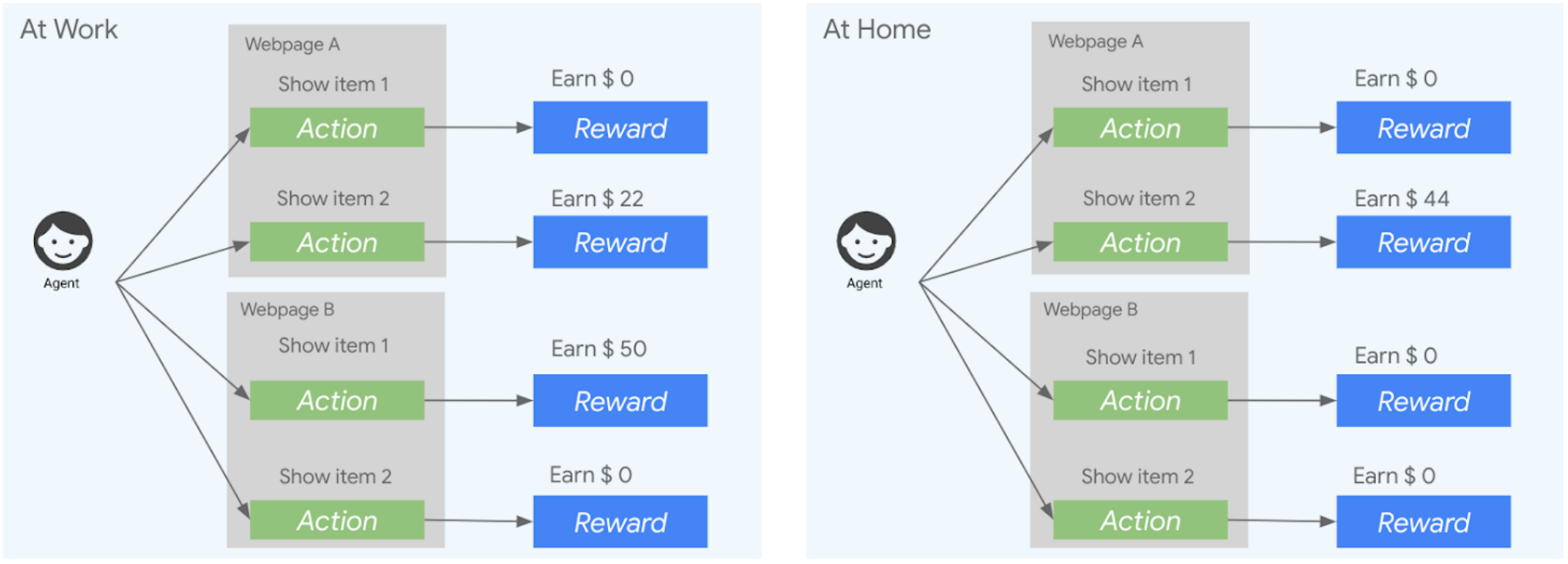
PROBLEM FORMULATION

Contextual Bandit

The learner and the environment interact in several rounds. In each round t,

- the leaner **observes a context** c_t from the environment
- the learner **chooses an action** $x_t \in A$, where A is action set
- the learner **receives a reward** y_t from the environment

The goal of the learner is to **maximize the cumulative reward** $\sum_{t=1}^{T} y_t$



In this work, we consider

- **Distributed** Stochastic Contextual Bandit
- Each agent cannot observe the context c_t rather only a **context distribution** μ_t
- The fixed and unknown stochastic linear reward function $y: A \times C \to \mathbb{R}$
  - $y_{t,i} := \langle \theta^*, \phi_{x_{t,i},c_t} \rangle + \eta_{t,i}$
  - $\phi_{x_{t,i},c_t} \in \mathbb{R}^d$ is a feature vector associated with context-action pair $(x_{t,i}, c_t)$
  - $\theta^* \in \mathbb{R}^d$ is the unknown reward parameter
  - $\eta_{t,t}$ is σ-subGaussian noise

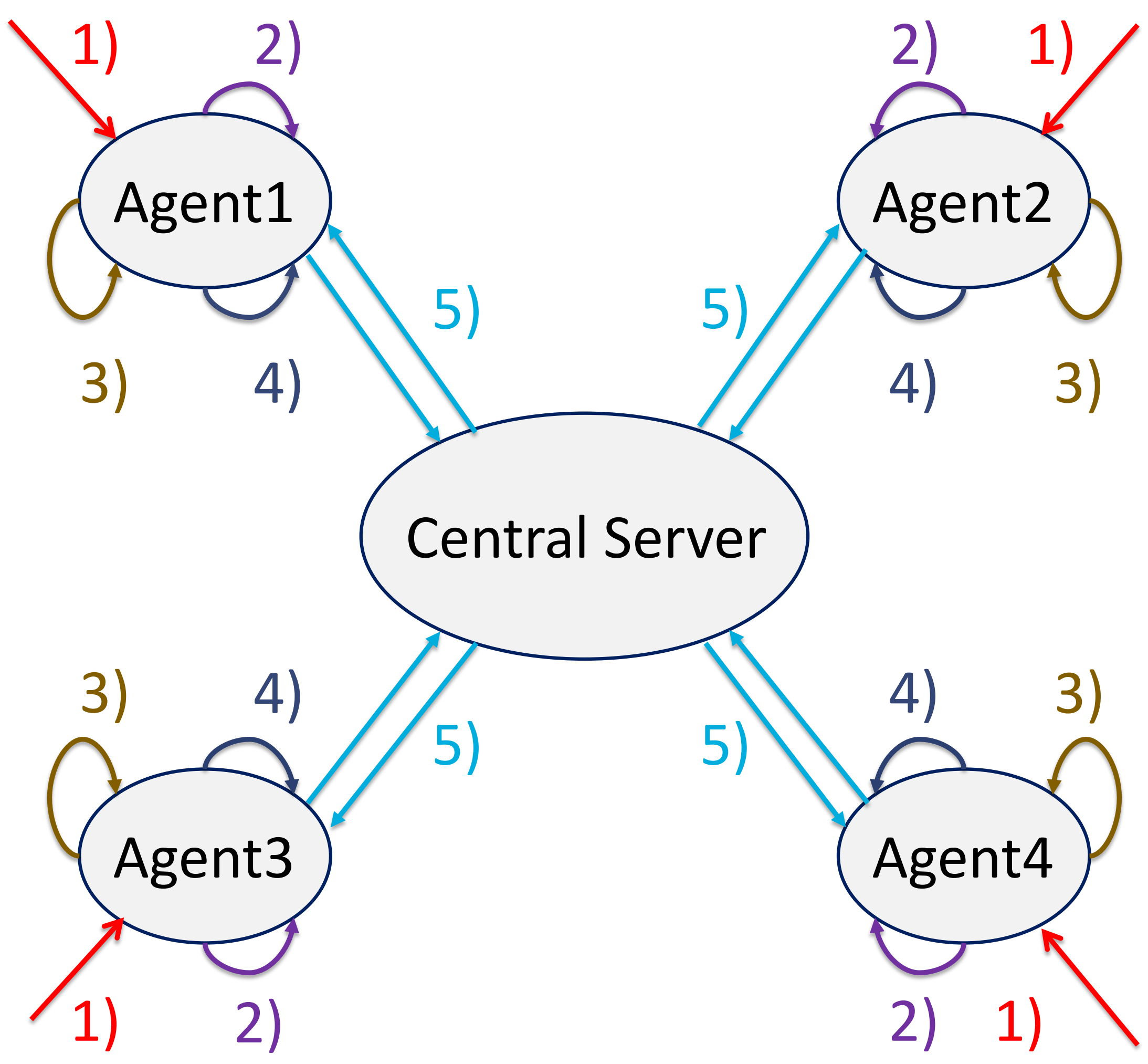
M agents jointly minimize cumulative regret

$$\mathcal{R}(T) = \sum_{i=1}^{M}\sum_{t=1}^{T}\left(\left\langle \theta^*, \phi_{x^*_{t,i},c_t}\right\rangle - \left\langle \theta^*, \phi_{x_{t,i},c_t}\right\rangle\right)$$

The **protein-drug interaction prediction** problem was modeled using bandit learning

PROPOSED APPROACH

Distributed UCB for LBs with hidden contexts



Key step of our Approach

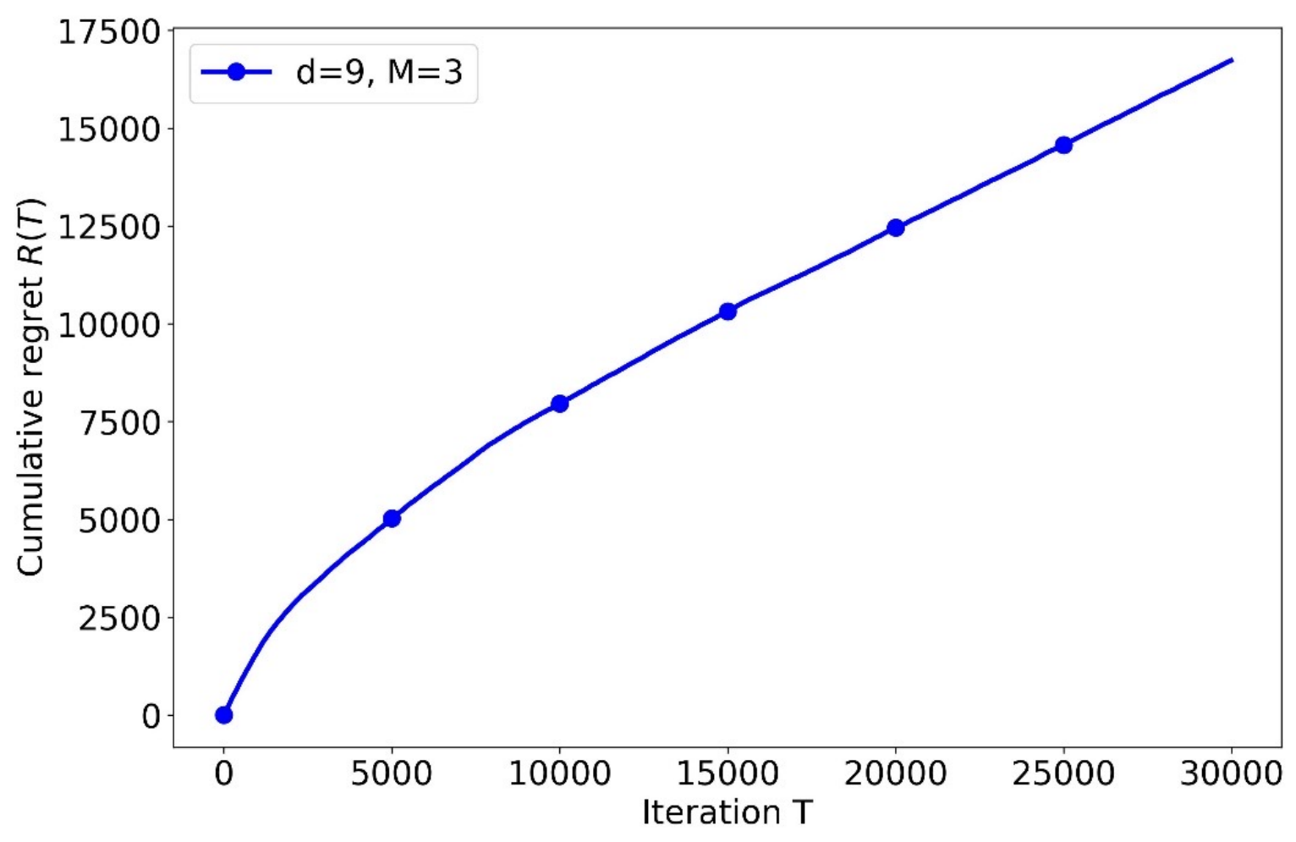
1) Each agent observes context distribution and construct feature vector set
2) Each agent chooses an optimistic estimate and action pair based on UCB (Upper Confidence Bound) approach
3) Each agent plays their respective optimistic actions and receive reward
4) Each agent uses the feature vector and reward to update their parameter
5) If the parameter change is significant,
   - All agents send their local estimates to central server
   - Central server computes the global parameters
   - Central server broadcasts the global parameters to each agents

EXPERIMENTAL RESULTS

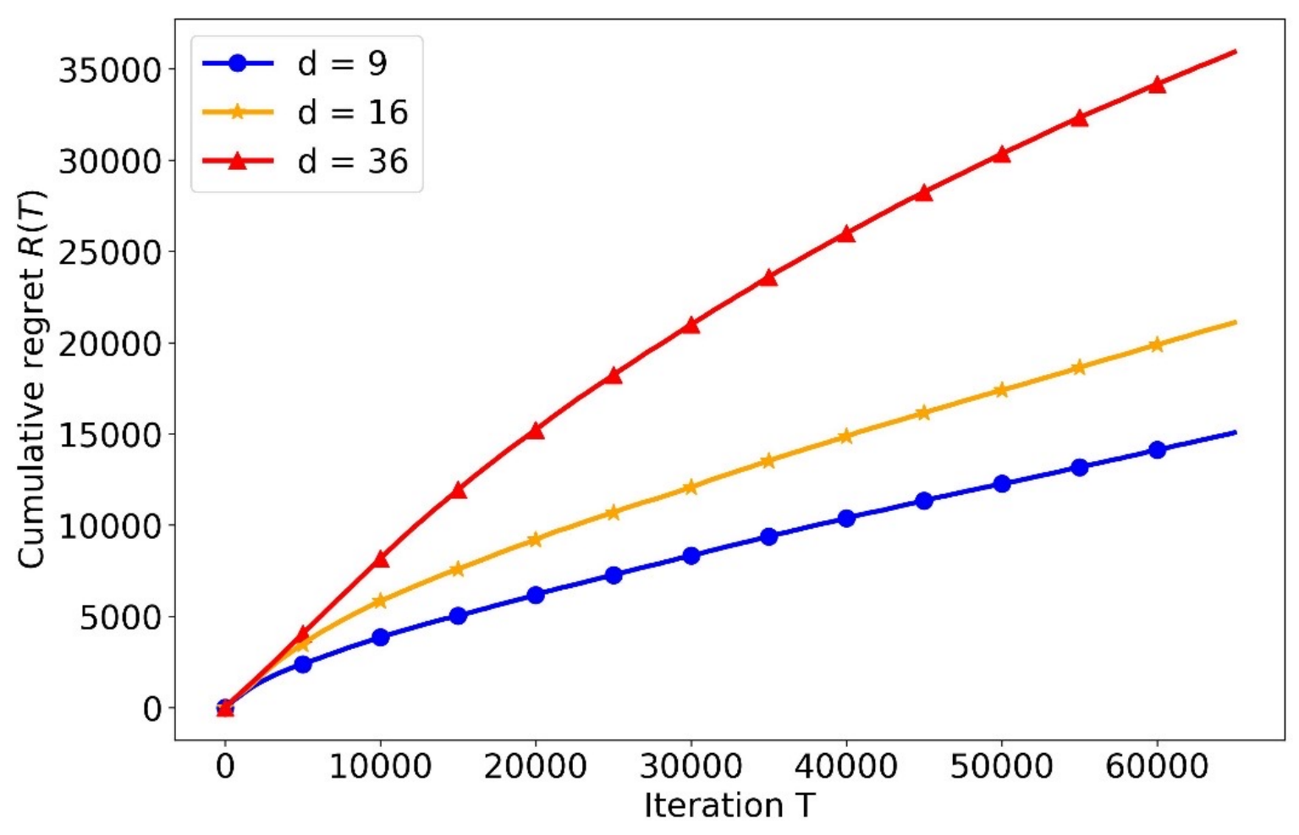
We generate latent representations using specific binding data from the **Harvard Medical School LINCS Center**

- Using NMF to decompose the data into context and action matrices
- Taking the 10-dimensional latent representing of each protein

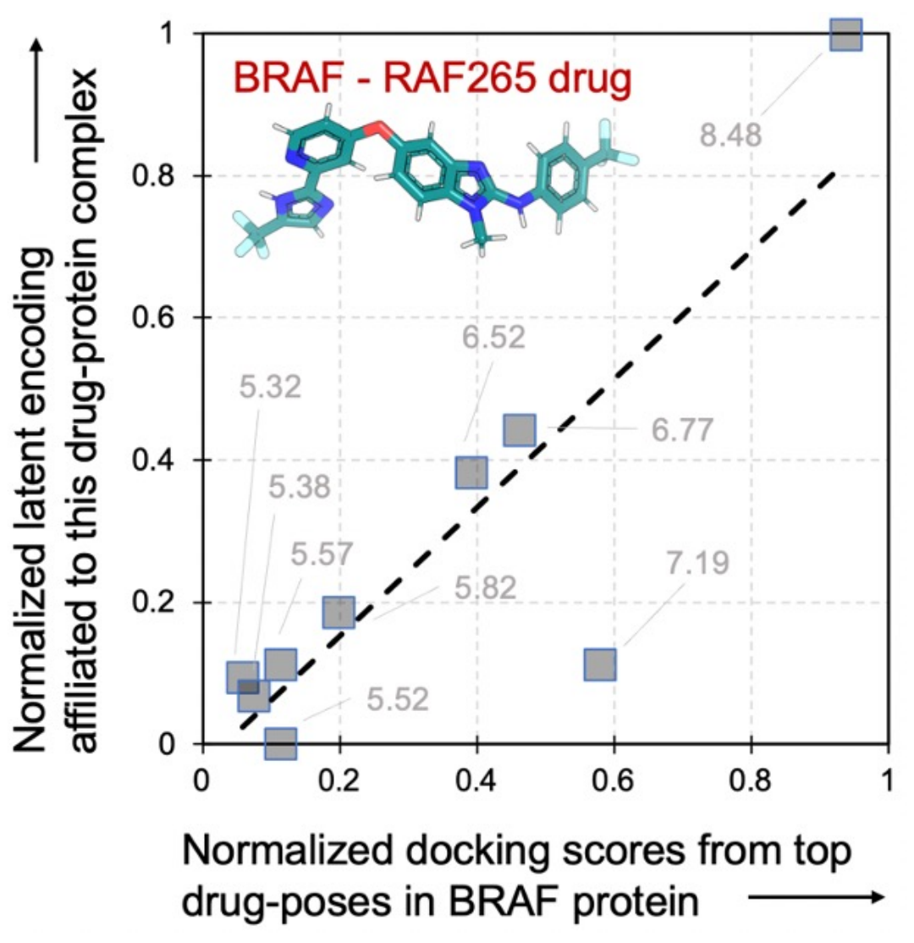
EXPERIMENTAL RESULTS

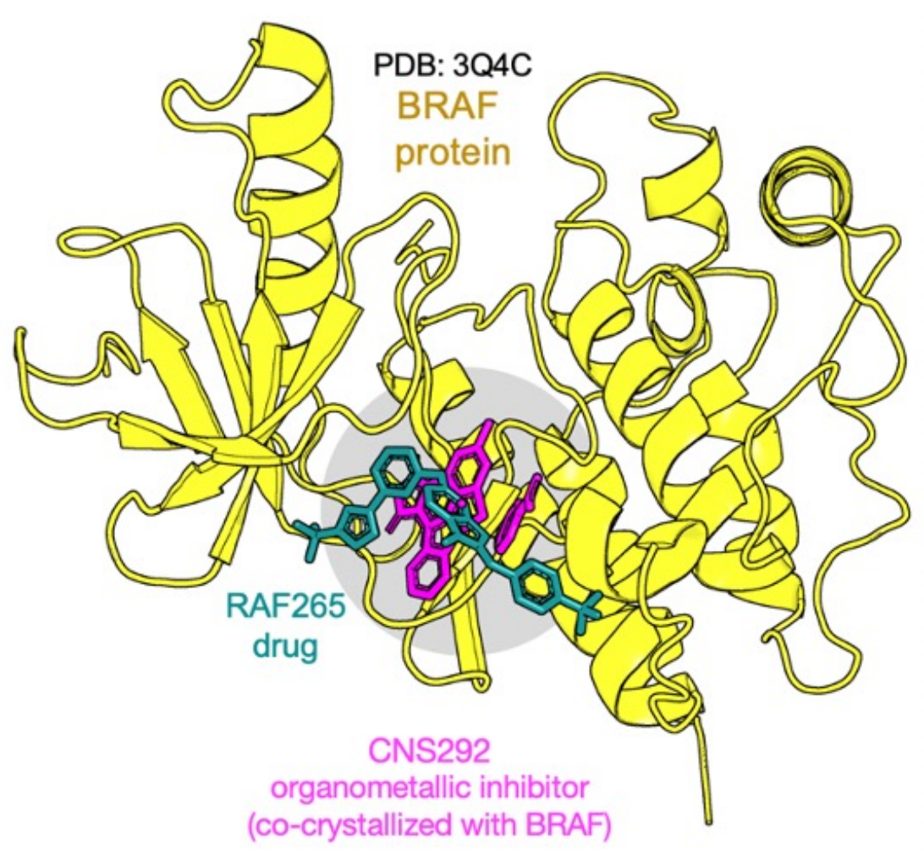


Cumulative regret is sublinear and gradually converges with increasing iterations.



Cumulative regret decreases as the dimension decreases.





The normalized docking energy scores and the normalized latent space representation of the drug-protein pair are highly correlated.

A representative of RAF265 that the drug bound to the known native drug-blinding pocket.

CONCLUSIONS

- Utilized specific protein-drug binding data to research the interaction properties through latent representations.
- Modeled the protein-drug interaction prediction as a bandit learning, aiming to learn binding relationships and select proteins for a given drug.
- Evaluated if the latent representations conform to any structural parameters that define this drug protein interactions