# Deep Hierarchy in Bandits

Joey Hong[1]
Branislav Kveton[2]
Sumeet Katariya[2]
Manzil Zaheer[3]
Mohammad Ghavamzadeh[4]

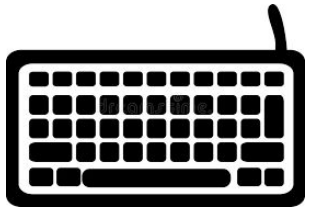
1 2 3 4



DeepMind

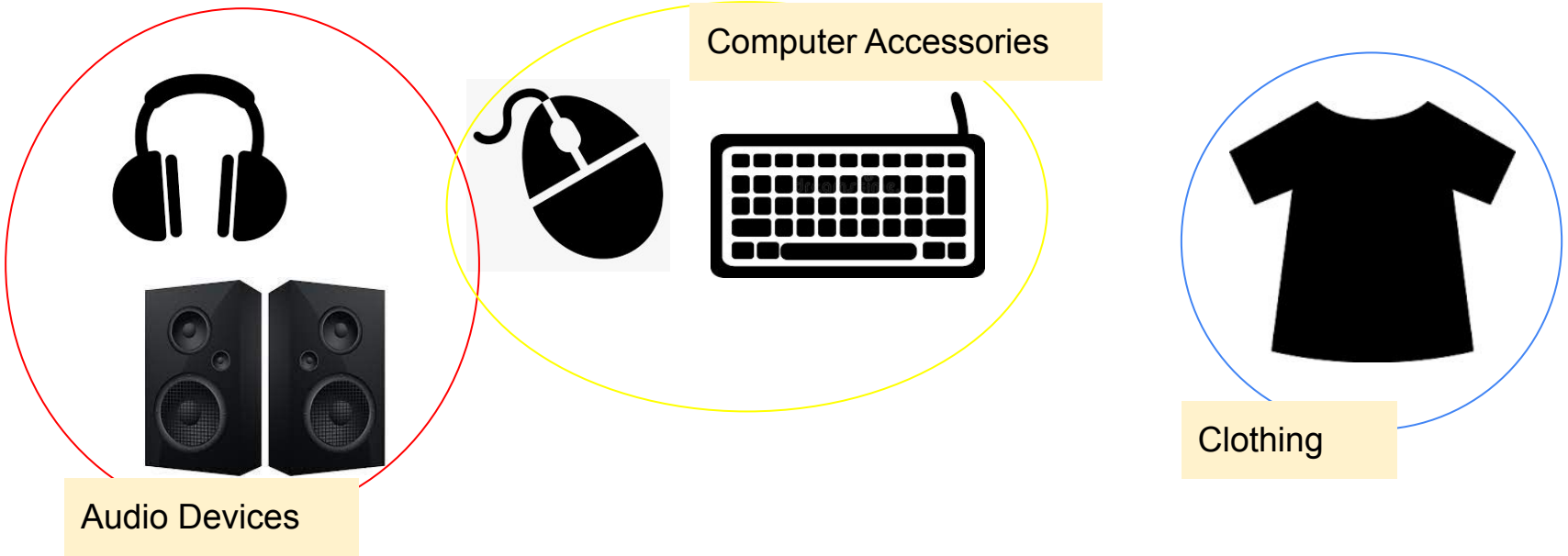
# Motivation: Online Shopping

- Goal: Given query, recommend best item from a large catalog



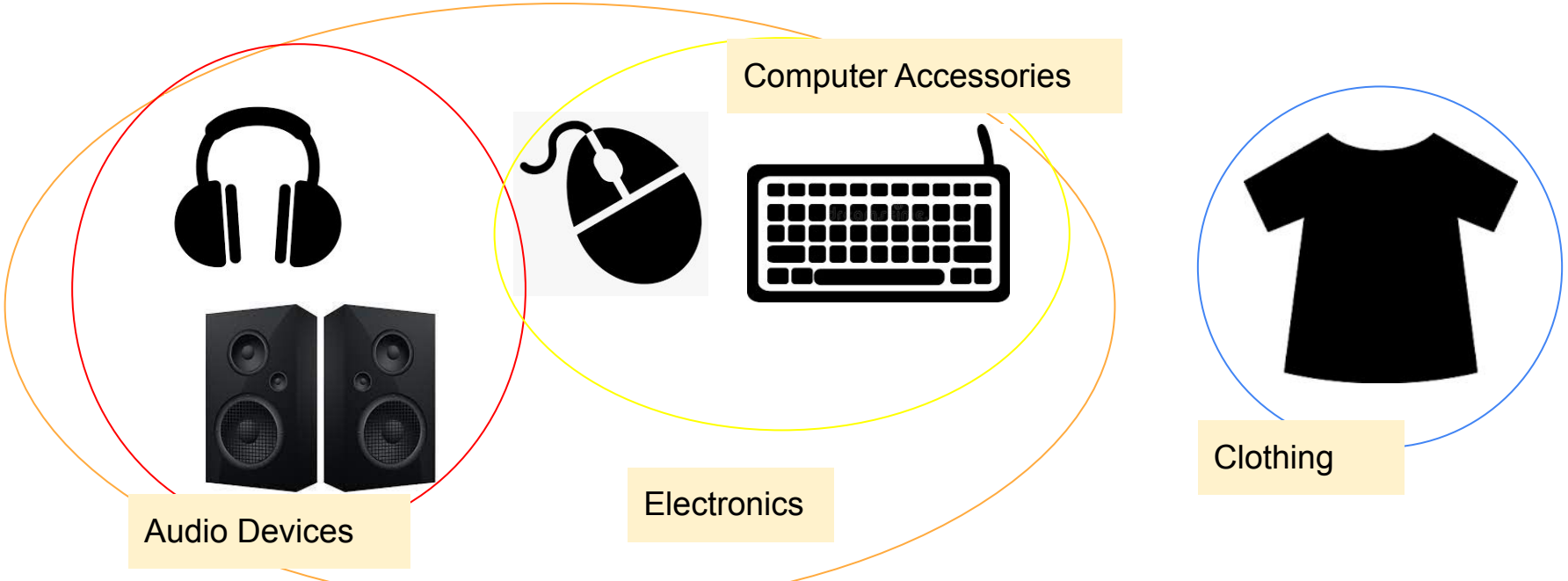
# Motivation: Online Shopping

- Goal: Given query, recommend best item from a large catalog
  - Assumption: Items are correlated with one another



# Motivation: Online Shopping

- Goal: Given query, recommend best item from a large catalog
  - Assumption: Items are correlated with one another



# Motivation

- Exploring one action can teach the system about other actions

# Motivation

- Exploring one action can teach the system about other actions

- Question 1: Can we improve upon naive exploration by leveraging the correlation between actions?

# Motivation

- Exploring one action can teach the system about other actions

- Question 1: Can we improve upon naive exploration by leveraging the correlation between actions?

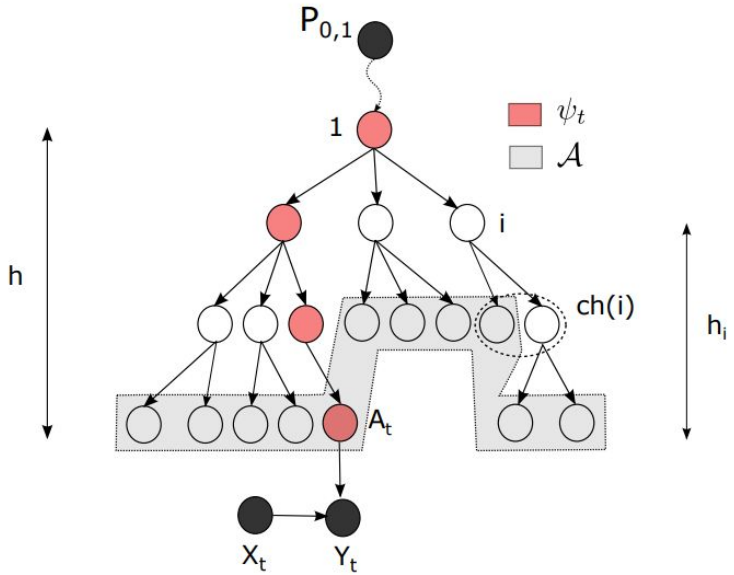
- Question 2: Can we do so in a computationally efficient manner?
  - Too slow: Naive Thompson sampling using joint posterior

# Setting

- Assumption: Action space can be represented as a hierarchical Bayesian model

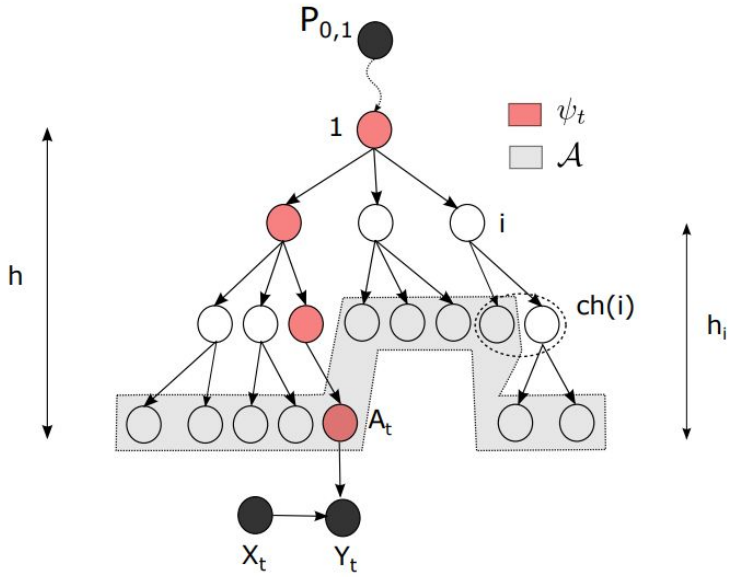
# Setting

- Assumption: Action space can be represented as a hierarchical Bayesian model
  - Actions are represented by action nodes/leaves



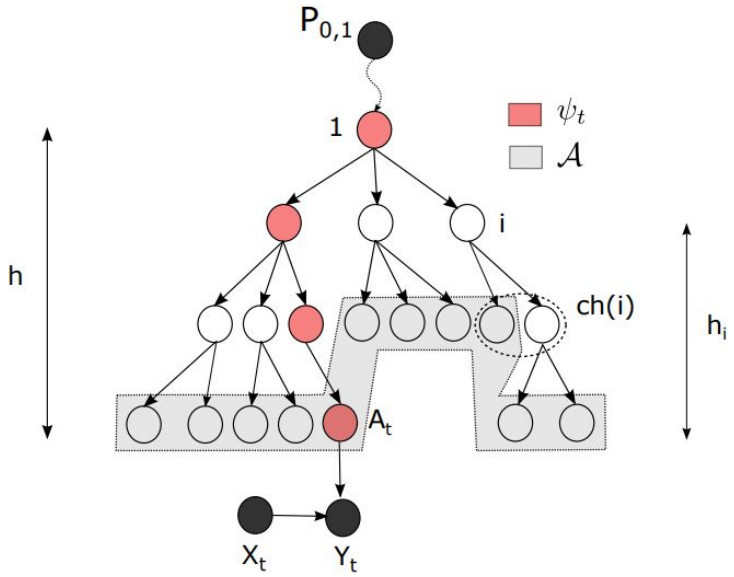
# Setting

- Assumption: Action space can be represented as a hierarchical Bayesian model
  - Actions are represented by action nodes/leaves
  - Correlations/clusters are internal nodes



# Setting

- Assumption: Action space can be represented as a hierarchical Bayesian model
  - Actions are represented by action nodes/leaves
  - Correlations/clusters are internal nodes
  - Each internal node has at most b children



# Hierarchical Bayesian Model

(Root node prior) $\theta_{*,1} \sim P_{0,1}$

# Hierarchical Bayesian Model

(Root node prior) $\theta_{*,1} \sim P_{0,1}$

(Conditional node prior) $\theta_{*,i} \mid \theta_{*,\mathsf{pa}(i)} \sim P_{0,i}(\cdot \mid \theta_{*,\mathsf{pa}(i)})$

Parent of node i

# Hierarchical Bayesian Model

(Root node prior) $\theta_{*,1} \sim P_{0,1}$

(Conditional node prior) $\theta_{*,i} \mid \theta_{*,\mathsf{pa}(i)} \sim P_{0,i}(\cdot \mid \theta_{*,\mathsf{pa}(i)})$

(Reward distribution) $Y_t \mid X_t, \theta_{*,A_t} \sim P(\cdot \mid X_t; \theta_{*,A_t}), \quad \forall t \in [n].$

Parent of node i

Only depends on action node

# Hierarchical Bayesian Model

(Root node prior) $\theta_{*,1} \sim P_{0,1}$

(Conditional node prior) $\theta_{*,i} \mid \theta_{*,\mathsf{pa}(i)} \sim P_{0,i}(\cdot \mid \theta_{*,\mathsf{pa}(i)})$

(Reward distribution) $Y_t \mid X_t, \theta_{*,A_t} \sim P(\cdot \mid X_t; \theta_{*,A_t}), \quad \forall t \in [n].$

Parent of node i

Only depends on action node

Any two nodes are related by lowest common ancestor, inducing complex correlations

# Hierarchical Thompson Sampling (HierTS)

Input: Tree T, priors $P_0$

Initialize $P_1 \leftarrow P_0$

For round t = 1, 2, …

Sample $\theta_{t,1} \sim P_{t,1}$

For node i in breadth-first traversal of T

Sample $\theta_{t,i} \sim P_{t,i}(\cdot \mid \theta_{t,\mathsf{pa}(i)})$

Take action $A_t \leftarrow \arg\max_{a \in \mathcal{A}} r(a, X_t; \theta_{t,a})$ and observe reward $Y_t$

Update posteriors $P_t$

# Hierarchical Thompson Sampling (HierTS)

Input: Tree T, priors $P_0$

Initialize $P_1 \leftarrow P_0$

For round t = 1, 2, …

Sample $\theta_{t,1} \sim P_{t,1}$

For node i in breadth-first traversal of T

Sample $\theta_{t,i} \sim P_{t,i}(\cdot \mid \theta_{t,\mathsf{pa}(i)})$

Take action $A_t \leftarrow \arg\max_{a \in \mathcal{A}} r(a, X_t; \theta_{t,a})$ and observe reward $Y_t$

Update posteriors $P_t$

$$P_{t,i}(\theta_i \mid \theta_{\mathsf{pa}(i)}) \propto P_{0,i}(\theta_i \mid \theta_{\mathsf{pa}(i)}) \prod_{j \in \mathsf{ch}(i)} \mathbb{P}(H_{t,j} \mid \theta_{*,i} = \theta_i)$$

Conditional posteriors can be updated recursively and efficiently using only nodes in path of $A_t$

Observations from actions in subtree of node j

# Regret Bound

Theorem 1: The Bayes regret of HierTS is bounded as

$$\mathcal{BR}(n) \leq \sqrt{2n\mathcal{G}(n)\log(1/\delta)} + \sqrt{2/\pi}\sigma_{\max}Kn\delta\,,$$

where $\mathcal{G}(n) = \sum_{i\in\mathcal{V}} c^{h_i} w_i$

Scales with the conditional prior variance of node i

# Regret Bound

Theorem 1: The Bayes regret of HierTS is bounded as

$$\mathcal{BR}(n) \leq \sqrt{2n\mathcal{G}(n)\log(1/\delta)} + \sqrt{2/\pi}\sigma_{\max}Kn\delta\,,$$

where $\mathcal{G}(n) = \sum_{i\in\mathcal{V}} c^{h_i} w_i$

Scales with the conditional prior variance of node i

Increases exponentially in height, but number of nodes also decreases exponentially in height.

# Regret Bound

- When conditional prior variances are equal:

  (Vanilla TS) $\mathcal{BR}(n) \leq \sqrt{2nhK \log(1/\delta)} + \sqrt{2/\pi}\sigma_{\max} K n\delta \,,$

# Regret Bound

- When conditional prior variances are equal:

(Vanilla TS) $$\mathcal{BR}(n) \leq \sqrt{2nhK \log(1/\delta)} + \sqrt{2/\pi}\sigma_{\max} K n\delta \,,$$

(HierTS) $$\mathcal{BR}(n) \leq \sqrt{2nK \log(1/\delta)} + \sqrt{2/\pi}\sigma_{\max} K n\delta \,,$$

# Regret Bound

- When conditional prior variances are equal:

  (Vanilla TS) $\mathcal{BR}(n) \leq \sqrt{2nhK \log(1/\delta)} + \sqrt{2/\pi}\sigma_{\max} K n\delta \,,$

  (HierTS) $\mathcal{BR}(n) \leq \sqrt{2nK \log(1/\delta)} + \sqrt{2/\pi}\sigma_{\max} K n\delta \,,$

- When conditional prior variances double with height:

  (Vanilla TS) $\mathcal{BR}(n) \leq \sqrt{2n2^h K \log(1/\delta)} + \sqrt{2/\pi}\sigma_{\max} K n\delta \,,$

# Regret Bound

- When conditional prior variances are equal:

  (Vanilla TS) $\mathcal{BR}(n) \leq \sqrt{2nhK \log(1/\delta)} + \sqrt{2/\pi}\sigma_{\max} K n\delta \,,$

  (HierTS) $\mathcal{BR}(n) \leq \sqrt{2nK \log(1/\delta)} + \sqrt{2/\pi}\sigma_{\max} K n\delta \,,$

- When conditional prior variances double with height:

  (Vanilla TS) $\mathcal{BR}(n) \leq \sqrt{2n2^h K \log(1/\delta)} + \sqrt{2/\pi}\sigma_{\max} K n\delta \,,$

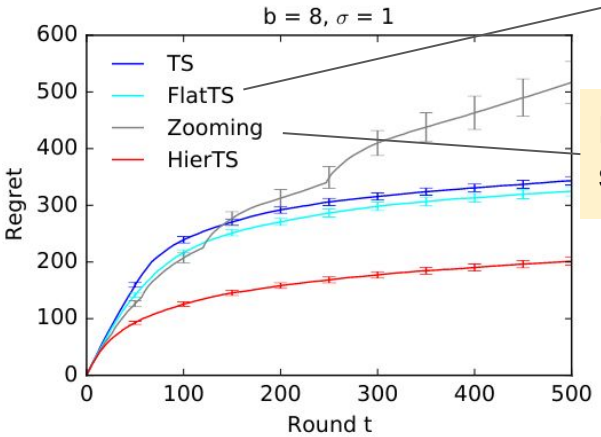
  (HierTS) $\mathcal{BR}(n) \leq \sqrt{2nK \log(1/\delta)} + \sqrt{2/\pi}\sigma_{\max} K n\delta \,,$

Since $2^h = K^{1/\log b}$, improvement is polynomial in K!

# Experiments

Evaluate on synthetic Gaussian multi-armed bandit with varying branching factor b, and height h

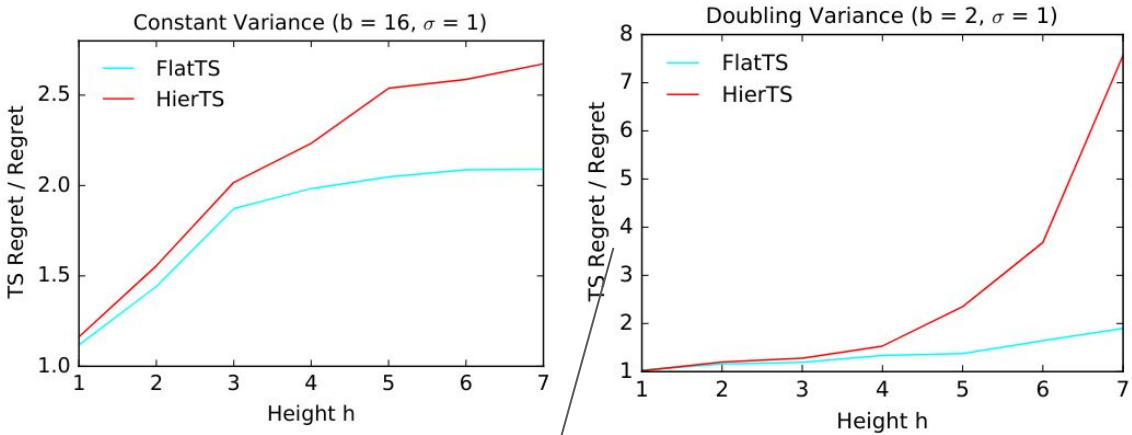
TS using last 2 levels of hierarchy




Existing UCB-like algorithm using metric space over actions

# Experiments

Evaluate on synthetic Gaussian multi-armed bandit with varying branching factor b, and height h




When variance doubles with level, HierTS improves exponentially in h over FlatTS

# Conclusions

- Study contextual bandit problem with hierarchical Bayesian model over actions

# Conclusions

- Study contextual bandit problem with hierarchical Bayesian model over actions
- Propose hierarchical Thompson sampling (HierTS)

# Conclusions

- Study contextual bandit problem with hierarchical Bayesian model over actions
- Propose hierarchical Thompson sampling (HierTS)
  - Efficient posterior updates using factorized joint posterior

# Conclusions

- Study contextual bandit problem with hierarchical Bayesian model over actions
- Propose hierarchical Thompson sampling (HierTS)
  - Efficient posterior updates using factorized joint posterior
  - Exact implementation using closed-form posteriors in Gaussian models

# Conclusions

- Study contextual bandit problem with hierarchical Bayesian model over actions
- Propose hierarchical Thompson sampling (HierTS)
  - Efficient posterior updates using factorized joint posterior
  - Exact implementation using closed-form posteriors in Gaussian models
  - Greatly-improved statistically efficient exploration over vanilla Thompson sampling