Thompson-Sampling-Based Wireless Transmission for Panoramic Video Streaming

Jiangong Chen†, Bin Li†, and R. Srikant*

†Dept. of ECBE
University of Rhode Island

*Dept. of ECE and CSL
University of Illinois at Urbana-Champaign
Motivation

• Emerging Commercial Head-mounted Displays (HMDs)

  Google Daydream  Facebook Oculus  Samsung Gear

• Panoramic video streaming provides immersive experience for users as if they are in a virtual 3D world

• **Main challenge:** it typically consumes 4~6x bandwidth of a regular video with the same resolution
Opportunity

- A user may only see as low as 20% of 360° scenes, known as **Field of View (FoV)**. It is sufficient to deliver 20% of 360° video scenes under **perfect motion prediction**.
Practical Challenges and Goal

- **Imperfect prediction**: should deliver a portion larger than the FoV
- **Time-varying wireless environment**: should quickly identify the optimal delivered portion
- There are a finite number of content portions covering the FoV and the goal is to quickly determine a portion with the maximum throughput.

Example of a set of content portions
Multi-armed Bandit Formulation

• Transmission fails for one of two reasons:
  • **FoV prediction**: If the selected portion covers the actual FoV, then the prediction is successful. Otherwise, the prediction fails.
  • **Wireless transmission**: If the rate of the selected portion is smaller than the channel rate, then the transmission is successful. Else, the transmission fails.

• Each arm $n$ corresponds to the selected portion or rate $r_n$. Each arm is associated with a success probability $\gamma_n$

• If all statistics are available, our goal is to select an arm satisfying

$$ n^* \in \text{argmax}_{n=1,2,...,N} r_n \gamma_n $$
MAB Formulation (Cont’d)

• However, statistics are unknown. Therefore, we need to dynamically select an arm with the goal of minimizing the regret.

\[
\text{Reg}(T) \triangleq r_n^* \gamma_n^* T - \mathbb{E} \left[ \sum_{t=1}^{T} r_{I(t)} Z_{I(t)}(t) \right]
\]

$I(t)$: the index of the selected rate in time slot $t$

$Z_n(t)$: indicates success or not in time slot $t$
Refined MAB Formulation

• After each play, we have both prediction and transmission outcomes of the user.
  • Even when the transmission fails, the HMD device automatically records the user’s orientation and sends back to the server for the next decision

• Each arm $n$ corresponds to the selected portion or rate $r_n$. Each arm is associated with a successful prediction probability $\alpha_n$ and a successful transmission probability $\beta_n$.

• If all statistics are available, our goal is to select an arm satisfying

$$n^* \in \arg\max_{n=1,2,...,N} r_n \alpha_n \beta_n$$
Refined MAB Formulation (Cont’d)

• As before, minimize regret

\[ \text{Reg}(T) \triangleq r_n^* \alpha_n^* \beta_n^* T - \mathbb{E} \left[ \sum_{t=1}^{T} r_{I(t)} X_{I(t)}(t) Y_{I(t)}(t) \right] \]

\( I(t) \): the index of the selected rate in time slot \( t \)

\( X_n(t) \): indicates whether the prediction is successful or not in time slot \( t \)

\( Y_n(t) \): indicates whether the transmission is successful or not in slot \( t \)
Standard KL UCB

- For each arm $n$, assign an index $\hat{\gamma}_n$ which is the largest value of $\gamma$ that satisfies
  \[ D(\hat{\gamma}_n(t) || \gamma) \leq \epsilon_n(t), \]
  where $\hat{\gamma}_n(t)$ is the empirical success probability at time $t$ and $\epsilon_n(t)$ is appropriately chosen
- Pull the arm with the largest index

- Extension to two-level feedback?
KL UCB for Two-Level Feedback

- Possibility 1: pick an index for the wireless part and the prediction part separately
  - $\max_\alpha D(\hat{\alpha}_n(t) || \alpha) \leq \epsilon_1 n(t)$, $\max_\beta D(\hat{\beta}_n(t) || \beta) \leq \epsilon_2 n(t)$
  - Index is $\tilde{\alpha}_n \ast \tilde{\beta}_n$

- Possibility 2: pick an index based on the overall success
  - $\max_\gamma D(\hat{\alpha}_n(t) \hat{\beta}_n(t) || \gamma) \leq \epsilon_n(t)$
  - Index is $\tilde{\gamma}_n$

- These approaches don’t seem to work well
Thompson Sampling with Single Feedback

• Selecting the rate according to the posterior probability:

\[
I(t) = \arg\max_{n \in \{1, 2, \ldots, N\}} r_n \gamma_n(t)
\]

Draw \( \gamma_n(t) \sim Beta(S_n + 1, F_n + 1) \)

\( Beta(a, b) \) is the beta distribution whose pdf is:

\[
p_{a,b}(x) \triangleq x^{a-1}(1 - x)^{b-1} \cdot \frac{\Gamma(a + b)}{\Gamma(a)\Gamma(b)}
\]

Counter of successes

Counter of failures

The Gamma function
Thompson Sampling with Two-Level Feedback

• Selecting the rate according to the approximate probability:

\[ I(t) = \arg\max_{n \in \{1,2,\ldots,N\}} r_n \alpha_n(t) \beta_n(t) \]

• Maintain a pair of counters for each outcome in each arm
• Draw probabilities from two independent Beta distributions
Simulations

<table>
<thead>
<tr>
<th>Rate $r_n$</th>
<th>arm1</th>
<th>arm2</th>
<th>arm3</th>
<th>arm4</th>
<th>arm5</th>
</tr>
</thead>
<tbody>
<tr>
<td>Prediction prob. $\alpha_n$</td>
<td>0.1</td>
<td>0.3</td>
<td>0.5</td>
<td>0.65</td>
<td>0.9</td>
</tr>
<tr>
<td>Transmission prob. $\beta_n$</td>
<td>0.99</td>
<td>0.6</td>
<td>0.4</td>
<td>0.2</td>
<td>0.05</td>
</tr>
<tr>
<td>Average throughput</td>
<td>0.198</td>
<td>0.54</td>
<td>1</td>
<td>0.78</td>
<td>0.405</td>
</tr>
</tbody>
</table>
Simulations (Cont’d)

• We used the dataset from [Bao, Wu, Zhang, Ramli, Liu, 2016] and predict the user’s orientation using linear regression.

• We simulated wireless transmission

• We got the estimated probabilities for each rate as follows by running experiments with fixed rate.
  
  $\alpha = [0.034, 0.708, 0.892, 0.990]$
  $\beta = [0.749, 0.599, 0.099, 0.030]$
Regret Lower Bound

• Lower bound for single feedback:

\[
\frac{r_1 \alpha_1 \beta_1 - r_2 \alpha_2 \beta_2}{D(\alpha_2 \beta_2 \| \alpha_1 \beta_1)} \log(t)
\]

• Lower bound for two-level feedback:

\[
\frac{r_1 \alpha_1 \beta_1 - r_2 \alpha_2 \beta_2}{D(\alpha_2 \| \alpha_1) + D(\beta_2 \| \beta_1)} \log(t)
\]
\( D(\alpha_1 || \alpha_2) + D(\beta_1 || \beta_2) \geq D(\alpha_1 \beta_1 || \alpha_2 \beta_2) \)?

- **Yes!** Consider independent random variables
  
  \[ X_1 \sim Ber(\alpha_1), Y_1 \sim Ber(\alpha_2), X_2 \sim Ber(\beta_1), Y_2 \sim Ber(\beta_2) \]

- Independence gives
  
  \[ D((X_1, Y_1)|| (X_2, Y_2)) = D(X_1, X_2) + D(Y_1, Y_2) = LHS \]

- **Data Processing Inequality:**
  
  \[ D((X_1, Y_1)|| (X_2, Y_2)) \geq D(X_1 Y_1 || X_2 Y_2) = RHS \]
Conclusions

• Formulated the problem of adaptive rate selection for panoramic video streaming as a multi-armed bandit problem with two-level feedback.
  • Proposed a modified Thompson Sampling algorithm efficiently leveraging the two-level feedback information.

• Ongoing work
  • Matching upper bound
  • Intuitively, the larger the selected rate, the higher the successful prediction probability and the lower the successful transmission probability, i.e.,

\[ r_1 < r_2 < \cdots < r_N \quad \alpha_1 < \alpha_2 < \cdots < \alpha_N \]
\[ \beta_1 > \beta_2 > \cdots > \beta_N \]
Related Work

• Panoramic Video Transmission
  • e.g., [Guan, Zheng, Zhang, Guo, Jiang, 2019], [Qian, Han, Xiao, Gopalakrishnan, 2018], [Bao, Wu, Zhang, Ramli, Liu, 2016]

• Reinforcement Learning Approach
  • e.g., [Zhang, Zhao, Bian, Liu, Song, Li, 2019], [Kan, Zou, Tang, Li, Liu, Xiong, 2019], [Xu, Song, Wang, Qiao, Huo, Wang, 2018]

• Multi-armed Bandit Problem
  • e.g., [Lattimore, Szepesvári, 2018], [Agrawal, Goyal, 2013], [Kaufmann, Korda, Munos, 2012]