Efficient Nonmyopic Active Search

Jiang, Malkomes, Converse, Shofner, Moseley and Garnett

STA 4273/CSC 2547 Paper Presentation

Presented by: Zain Hasan & Daniel Hidru
Active Search

- Sequentially locating as many members of a particular class as possible - targets that belong to a rare class

\[ \mathcal{D} \triangleq \{(x_i, y_i)\} \quad y \triangleq 1\{x \in \mathcal{R}\} \]

- Active search is Bayesian optimization with binary rewards and cumulative regret (budget).
Analogy for Active Search

- Writing a Literature Review is an active search process
  - Limited amount of papers you can read (budget)
  - Reading papers you know are relevant (exploitation)
  - Reading papers that might be relevant in the hope that you find more relevant papers (exploration)
Budget (Cumulative regret)

- You have limited time (deadline) and resources.
- Have to balance between exploration and exploitation to maximize utility for binary $y = \{0, 1\}$:

\[
u(D) \triangleq \sum_{y_i \in D} y_i,
\]

- Which counts the number of targets in chosen set (i.e. Relevant papers included in review)
- Want to determine/approximate some optimal policy of picking points that maximizes utility
Myopic vs. Nonmyopic

- Myopic search: consider the effect of only potential immediate choices
  - Easier, lower runtime complexity, but short-sighted

- Nonmyopic search: consider impact of all selected points, immediate and future
  - Harder, more complex, but potentially better results
Contributions of Paper

1. Prove that active search, that approximates the optimal policy, is hard to do by finding its runtime complexity via a proof

2. Suggest an efficient nonmyopic search algorithm
Background for Algorithm

- Optimal Bayesian Decision/Policy:
  - Posterior prob. of a point belonging to desired $y = 1$ class
    \[
    \Pr(y = 1 \mid x, D)
    \]

- Choose next points maximizing the expected number of targets found at termination, given $i - 1$ previous observations:
  \[
  x_i^* = \arg \max_{x_i \in \mathcal{X} \setminus D_{i-1}} \mathbb{E}[u(D_t) \mid x_i, D_{i-1}]
  \]
Expected utility: 1 query left

\[
\mathbb{E}[u(D_t) \mid x_t, D_{t-1}] = \sum_{y_t} u(D_t) \Pr(y_t \mid x_t, D_{t-1})
\]

\[
= u(D_{t-1}) + \Pr(y_t = 1 \mid x_t, D_{t-1}). \quad (3)
\]

Expected utility of selecting \(x_t\), given previous selections \((D_{t-1}) =

Reward for previous selections + Expected reward of current selection

- Pure exploitation because there are no more queries to make
Expected utility: 2 queries left

\[
\mathbb{E}[u(D_t) \mid x_{t-1}, D_{t-2}] = u(D_{t-2}) + \\
\Pr(y_{t-1} = 1 \mid x_{t-1}, D_{t-2}) + \\
\mathbb{E}_{y_{t-1}} \left[ \max_{x_t} \Pr(y_t = 1 \mid x_t, D_{t-1}) \right].
\]

Expected utility of selecting \(x_{t-1}\), given previous selections \((D_{t-2}) = \)

Reward for previous selections + Expected reward of current selection + Expected reward for final selection given outcome of current selection

- Natural trade off between exploitation (2nd term) and exploration (3rd term)
Expected utility: t-i+1 queries left

\[ E[u(D_t) \mid x_i, D_{i-1}] = u(D_{i-1}) + \]
\[ \Pr(y_i = 1 \mid x_i, D_{i-1}) + \]
\[ \underbrace{E_{y_i} \left[ \max_{x'} E \left[ u(D_t \setminus D_i) \mid x', D_i \right] \right]}_{\text{exploration, } < t-i} \]

Expected utility of selecting \( x_i \), given previous selections (\( D_{i-1} \)) =

Reward for previous selections + Expected reward of current selection + Expected reward for remaining selections given outcome of current selection

- Can compute this expectation recursively
  - Cost: exponential in the number of future queries - \( O((2n)^l) \)
Theorem 1. There is no polynomial-time active search policy with a constant factor approximation ratio for optimizing the expected utility.
Efficient Nonmyopic Search (ENS): t-i+1 queries left

\[
\max_{x'} \mathbb{E}[u(D_t \setminus D_i) \mid x', D_i] \approx \sum_{t-i} \Pr(y = 1 \mid x, D_i),
\]

\[
\mathbb{E}[u(D_t) \mid x_i, D_{i-1}] \approx u(D_{i-1}) + \Pr(y_i = 1 \mid x_i, D_{i-1}) + \mathbb{E}_{y_i} \left[ \sum_{t-i} \Pr(y = 1 \mid x, D_i) \right]
\]

Expected utility of selecting \(x_i\), given previous selections \((D_{i-1})\) ≈

Reward for previous selections + Expected reward of current selection + Expected reward for remaining selections given they are selected as a batch

- Assumption: the labels of all unlabeled points are conditionally independent
  - Needed to reduce the final term to a sum of marginal probabilities
Assumptions for efficiency improvements

1. Updating the model only affects a limited number of samples.
2. Observing a new negative point will not raise the probability of any other point being a target.
3. Able to bound the maximum probability of the unlabeled data conditioned on the future selection of additional targets.
Representative experiment: CiteSeer data

- **Data:**
  - 39,788 computer science papers published in the top 50 venues
  - 2,190 (5.5%) are NIPS publications

- **Goal:** Find the most NIPS publications given a budget $t=500$

- **Model:** $k$-NN with $k=50$
  - Easy to update
  - Consistent with efficiency assumptions

- **Features:** graph PCA on the citation network using the first 20 principal components
Results: All 500 queries

Results: First 80 queries

Results: Different budgets

<table>
<thead>
<tr>
<th>policy</th>
<th>100</th>
<th>300</th>
<th>500</th>
<th>700</th>
<th>900</th>
</tr>
</thead>
<tbody>
<tr>
<td>one-step</td>
<td>25.5</td>
<td>80.5</td>
<td>141</td>
<td>209</td>
<td>273</td>
</tr>
<tr>
<td>two-step</td>
<td>24.9</td>
<td>89.8</td>
<td>155</td>
<td>220</td>
<td>287</td>
</tr>
<tr>
<td>ENS–900</td>
<td>25.9</td>
<td>94.3</td>
<td>163</td>
<td>239</td>
<td><strong>308</strong></td>
</tr>
<tr>
<td>ENS–700</td>
<td>28.0</td>
<td>105</td>
<td>188</td>
<td><strong>259</strong></td>
<td></td>
</tr>
<tr>
<td>ENS–500</td>
<td>28.7</td>
<td><strong>112</strong></td>
<td><strong>189</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ENS–300</td>
<td>26.4</td>
<td>105</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ENS–100</td>
<td><strong>30.7</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Relationship to other fields of research

- Active learning: train a high performing model with a few selected examples
  - AS: find elements of a rare class with a few selected choices
- Multi-armed bandit: maximize expected score given limited resources
  - AS: items are correlated and can only be selected once
  - ENS similar to knowledge gradient policy (Frazier et al., 2008)
- Bayesian optimization: global optimization using sequential choices
  - AS: special case with binary observations and cumulative reward
  - ENS similar to GLASSES algorithm (González et al., 2016)
Limitations (related to this course) and future work

- **Active Search/ENS Approach**
  - Can’t select the same element multiple times
    - Difficult to apply to reinforcement learning where the same action can be repeated
  - Can’t work in a continuous object domain
    - Needs discrete objects that can’t be selected multiple times to avoid selecting objects that are arbitrarily close to a previously selected item
  - True reward does not depend on previous actions
    - The order of the decisions affects your performance in reinforcement learning

- **Bayesian Optimization**
  - Probability models need to be updated multiple times before each selection
    - Costly to retrain neural networks (idea: update with a few gradient steps)
  - Difficult to work with continuous labels/rewards
    - Challenging to integrate the expected future reward (idea: estimate expectation)
Summary

- Efficient Nonmyopic Search outperforms myopic search in the active search problem by considering the benefit of exploration associated with the rewards of future queries.
- The key idea will be difficult to utilize in our course projects because it depends on many of the constraints imposed by the problem definition.
References