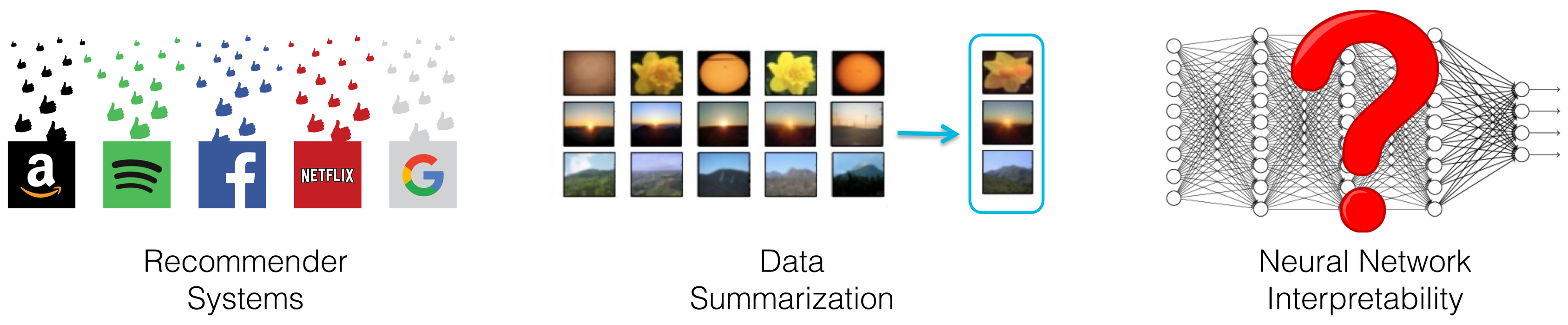
Adaptive Sequence Submodularity

Marko Mitrovic[1], Ehsan Kazemi[1], Moran Feldman[2], Andreas Krause[3] and Amin Karbasi[1]
[1]Yale University, [2]The Open University of Israel and [3]ETH Zurich

1. Background

- In a nutshell, submodular functions are the class of functions that exhibit diminishing returns. As such, many machine learning applications fall under the umbrella of submodularity:



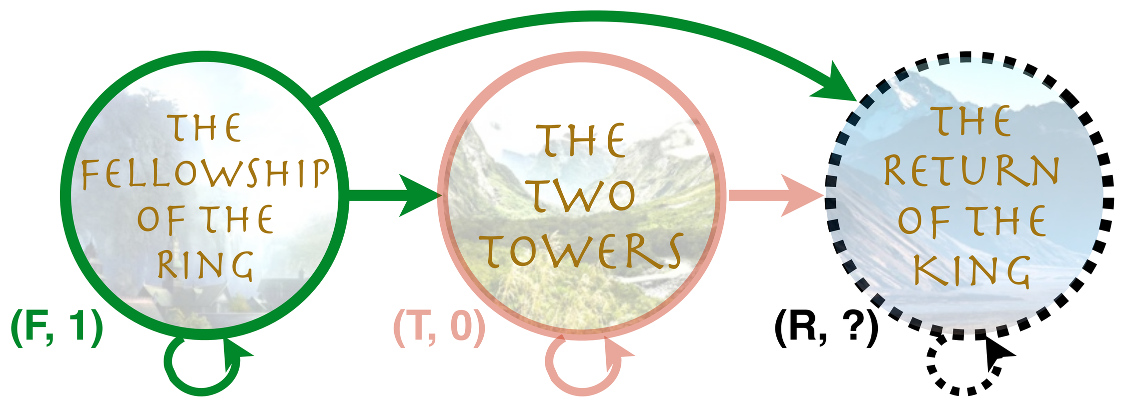
- Mathematically, a function is said to be **submodular** if for all sets $A \subseteq B$ and all elements $v \in V \backslash B$:

$$f(A \cup \{v\}) - f(A) \geq f(B \cup \{v\}) - f(B)$$

- In other words, the marginal value of any item is non-increasing as our set grows. For example, suppose we want to summarize a set of images about Vancouver. Once we already have one image of the convention centre, additional images of the convention centre will be much less valuable.

2. Problem Statement

- We view the problem of adaptive and sequential decision making through the lens of submodularity.
- We assume there is a directed graph $G = (V, E)$, where each item in our ground set is represented as a vertex $v \in V$, and the edges encode the additional value intrinsic to picking certain items in certain orders.
- A sequence of items σ induces a set of edges $E(\sigma) = \{(\sigma_i, \sigma_j) \mid (\sigma_i, \sigma_j) \in E, i \leq j\}$
- To include adaptivity, we assume that each vertex has some (initially unknown) state $o \in O$, each edge has a state $q \in Q$ that is determined entirely by the states of its endpoints. Another way to look at this is to say a realization ϕ of the vertices induces a realization ϕ^E of the edges.
- This allows us to define $f(\sigma, \phi) = h(E(\sigma), \phi^E)$, where h is a weakly-adaptive set submodular function, σ induces $E(\sigma)$, and ϕ induces ϕ^E



  - This example gives a possible partial realization of the vertices ψ_1 and an associated partial realization of the edges ψ_1^E. In this case, the state of an edge is equal to the state of its start point.
  - Suppose our function h counts all induced edges that are in state 1. Furthermore, let us simply assume that any unknown vertex is equally likely to be in state 0 or state 1. This means that the self-loop (R, R) is also equally likely to be in either state 0 or state 1. Therefore, $\Delta((R, R) \mid \psi_1^E) = \frac{1}{2} \times 0 + \frac{1}{2} \times 1 = \frac{1}{2}$.

3. Algorithm and Theoretical Results

Algorithm 1 Adaptive Sequence Greedy Policy π

Input: Directed graph $G = (V, E)$, weakly adaptive sequence submodular $f(\sigma, \phi) = h(E(\sigma), \phi^E)$, and cardinality constraint k
Let $\sigma \leftarrow ()$
while $|\sigma| \leq k - 2$ **do**
  $\mathcal{E} = \{e_{ij} \in E \mid v_j \notin \sigma\}$
  if $\mathcal{E} \neq \varnothing$ **then**
    $e_{ij} = \arg\max_{e \in \mathcal{E}} \Delta(e \mid \psi_\sigma^E)$
    if $v_i = v_j$ **or** $v_i \in \sigma$ **then**
      $\sigma = \sigma \oplus v_j$ and observe state of v_j
    else
      $\sigma = \sigma \oplus v_i \oplus v_j$ and observe states of v_i, v_j
    end if
  else
    break
  end if
end while
Return σ

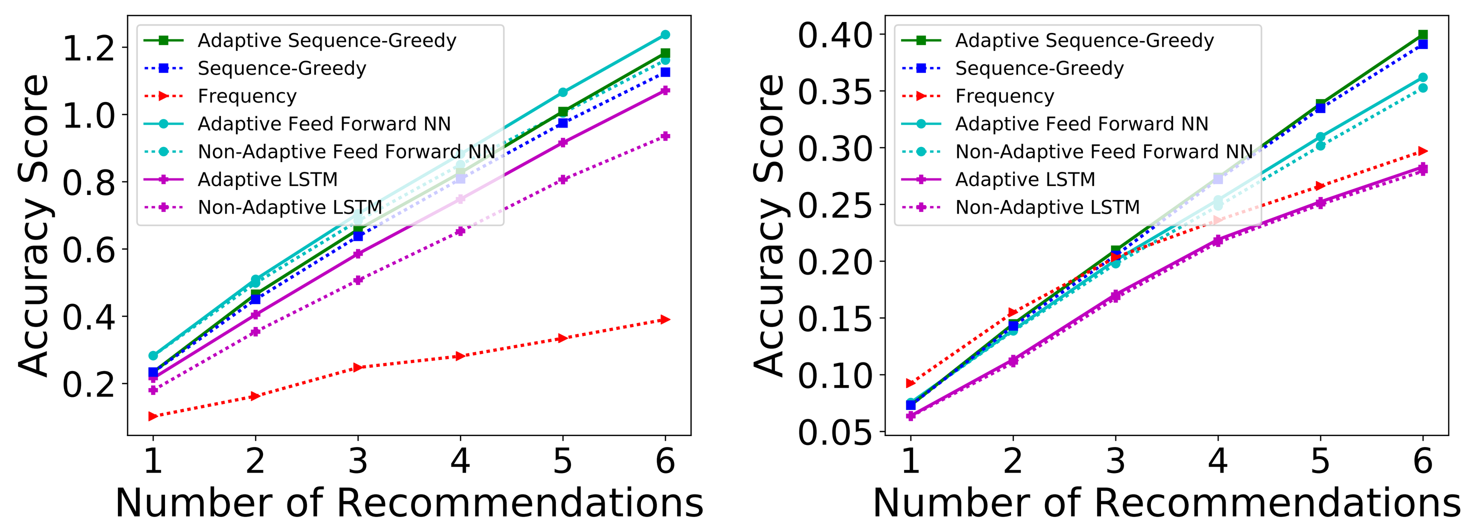
Theorem 1. *For adaptive monotone and weakly adaptive sequence submodular function f, the Adaptive Sequence Greedy policy π represented by Algorithm 1 achieves*

$$f_{avg}(\pi) \geq \frac{\gamma}{2d_{\text{in}} + \gamma} \cdot f_{avg}(\pi^*),$$

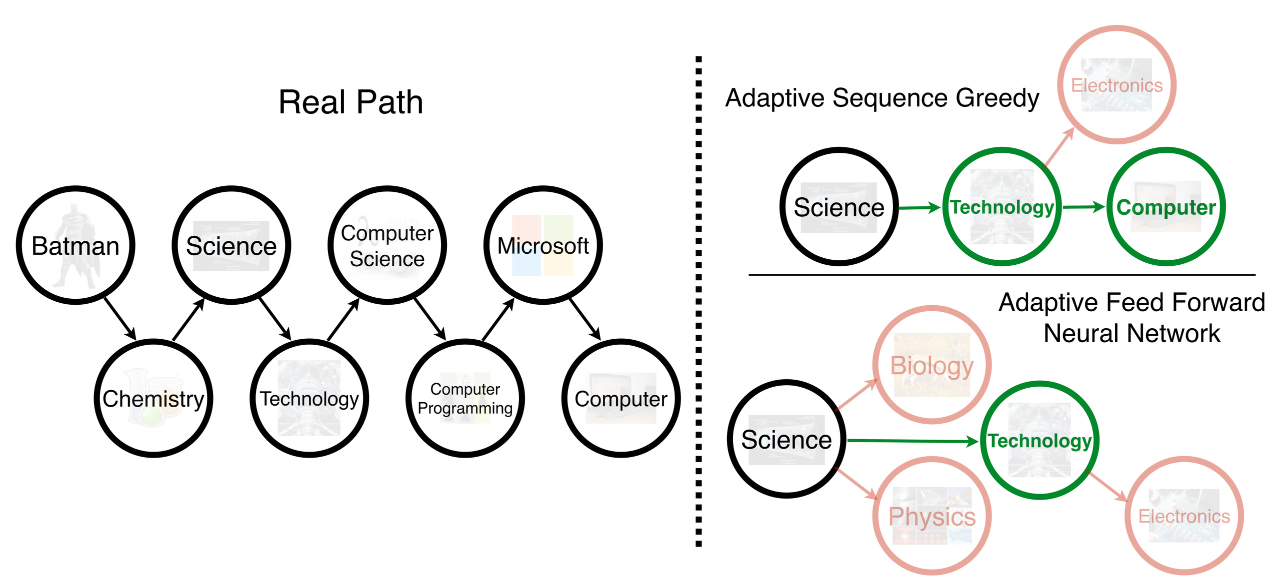
where γ is the weakly adaptive submodularity parameter, π^* is the policy with the highest expected value and d_{in} is the largest in-degree of the input graph G.

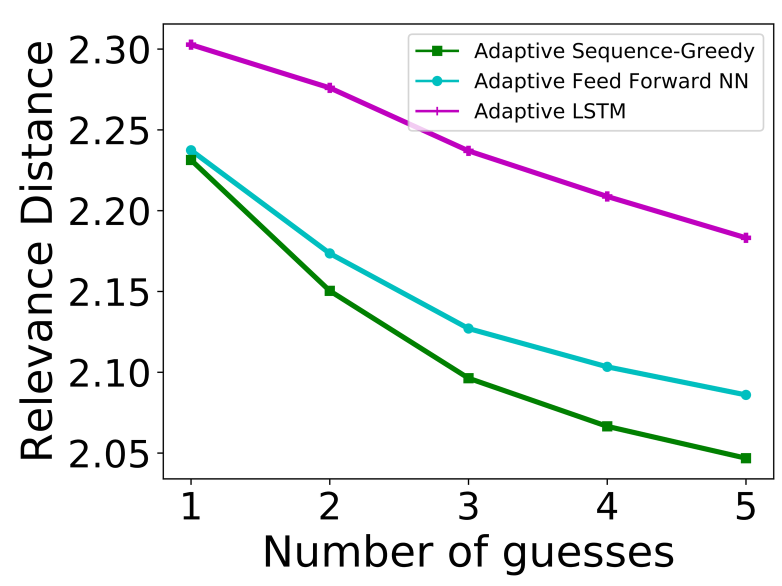
4. Applications

- Product Recommendation
  - We use the Amazon Video Games dataset (McAuley et al., 2015), which contains 10,672 products, 24,303 users, and 231,780 confirmed purchases.
  - Given the first 4 products that a user has purchased, our goal is to recommend k products that we think she will purchase.



- Wikipedia Link Prediction
  - We use the Wikispeedia dataset (West et al., 2009), which consists of 51,138 completed search paths on a condensed version of Wikipedia that contains 4,604 pages and 119,882 links between them.
  - Given the first 3 pages a user has visited, we want to guide her to her target page.





5. Acknowledgements

This work is partially supported by NSF (IIS-1845032), ONR (N00014- 19-1-2406), AFOSR (FA9550-18-1-0160), ISF (1357/16), and ERC StG SCADAPT.