Adaptive Sequence Submodularity

Marko Mitrovic¹ Ehsan Kazemi¹ Moran Feldman² Andreas Krause³ and Amin Karbasi¹

¹Yale University, ²The Open University of Israel and ³ETH 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 \( v \in V \setminus B \):

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

▶ 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 \( \sigma \) 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 \( \phi \) of the vertices induces a realization \( \phi^E \) of the edges.

▶ This allows us to define \( f(\sigma, \phi) = h_i(E(\sigma), \phi^E) \), where \( h \) is a weakly-adaptive set submodular function, \( \sigma \) induces \( E(\sigma) \), and \( \phi \) induces \( \phi^E \).

This example gives a possible partial realization of the vertices \( \psi \) and an associated partial realization of the edges \( \psi^F \). 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, \( \Delta(\{(R, R) \mid \psi^F\}) = \frac{1}{2} \times 0 + \frac{1}{2} \times 1 = \frac{1}{2} \).

3. Algorithm and Theoretical Results

Algorithm 1 Adaptive Sequence Greedy Policy \( \pi \)

1. Input: Directed graph \( G = (V, E) \), weakly adaptive sequence submodular \( f(\sigma, \phi) = M(\sigma(\sigma), \phi^E) \), and cardinality constraint \( \kappa \)
2. Let \( \pi = [\{\}] \)
3. While \( |\pi| < \kappa \) do
   4. \( E = \{e \mid e \in E \mid \psi_e \neq \pi\} \)
   5. If \( E = \emptyset \) then
      6. \( \pi = \pi \cup \{a\} \)
     7. \( \pi = \pi \cup \{a\} \)
   8. Else
      9. \( \sigma = \sigma(\pi) \) and \( \phi = \phi(\pi) \)
     10. \( \pi = \pi \cup \{a\} \)
     11. \( \pi = \pi \cup \{a\} \)
   12. End if
   13. Break
   14. End if
16. End while
16. Return \( \pi \)

Theorem 1. For adaptive monotone and weakly adaptive sequence submodular function \( f \), the Adaptive Sequence Greedy policy \( \pi \) represented by Algorithm 1 achieves

\[
 I_{\text{act}}(\pi) \geq \frac{1}{2 \delta_{\text{min}} + \gamma} I_{\text{opt}}(\pi),
\]

where \( \gamma \) is the weakly adaptive submodularity parameter, \( \sigma^* \) is the policy with the highest expected value and \( \delta_{\text{min}} \) is the largest in-degree of the input graph \( G \).

4. Applications

▶ Product Recommendation

▶ We use the Amazon Video Games dataset (McCaulay 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 Wikispeeda 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 (1845032), ONR (N00014-19-1-2406), AFOSR (FA9550-18-1-0160), ISF (1305/16), and ERC StG SCADAPT.