TWK Literature Reviews for Sparse Supermod

Taylor Killian

March 8, 2016

1 Boutsidis, et. al. 2015

What is the problem?

- Definition of *weak-α-supermodularity* for set functions
- Evaulate a greedy algorithm based on this definition in determining a minimal solution

Submodular function:

$$f(S \cup T) - f(S) \geq f(P \cup T) - f(P), \qquad \text{for } S \subseteq P$$

Supermodular beginnings:

$$f(S \cup T) \leq f(S) \Rightarrow f(S) - f(S \cup T) \text{ is the reduction in } f \text{ by adding } T$$

Then the average gain of adding elements of T *sequentially* is $\frac{[f(S)-f(S \cup T)]}{|T \setminus S|}$. One hopes that there exists an element $i \in T \setminus S$ such that $f(S) - f(S \cup \{i\}) \geq \frac{[f(S)-f(S \cup T)]}{|T \setminus S|}$. In general this is not possible since the elements of T are not independent. This is, however a feature of supermodular functions (Lemma 1).

Supermodular definition: *A set function $f(S) : 2^{[n]} \to \mathbb{R}_+$ is said to be supermodular if for any two sets $S, T \subseteq [n]$*

$$f(S \cap T) + f(S \cup T) \geq f(S) + f(T)$$

Extends to definition of *weak-α-supermodularity.* *A non-negative, non-increasing, set function $f(S) : 2^{[n]} \to \mathbb{R}_+$ is said to be weakly-α-supermodular if there exists $\alpha \geq 1$ such that for any two sets $S, T \in [n]$:*

$$f(S) - f(S \cup T) \leq \alpha |T \setminus S| \max_{i \in T \setminus S} [f(S) - f(S \cup \{i\})]$$

Why is it important?

- Provides a conversant analog to the submodular formulation in a few select problems. (Most particularly that of Sparse Multiple Linear Regression (SMLR))
- (In general, I have no idea why supermodularity is an improvement over submodularity...)

Why is it hard?

- Sparse combinatorial optimization based on cardinality constraints, hard to minimize in general

Why existing solutions do not work?

- Not all problems that are related through submodularity can be redefined as supermodular
- The average gain of adding elements of T *sequentially* is $\frac{[f(S)-f(S \cup T)]}{|T \setminus S|}$. One hopes that there exists an element $i \in T \setminus S$ such that $f(S) - f(S \cup \{i\}) \geq \frac{[f(S)-f(S \cup T)]}{|T \setminus S|}$. In general this is not possible since the elements of T are not independent. This is, however a feature of supermodular functions (Lemma 1).

What is the core intuition for the solution?

- That a non-increasing supermodular function is *weakly-α-supermodular* with $\alpha = 1$
- Many of the underlying results depend on those of referenced papers, intuition is built through the literature

Solution step-by-step?

- Iteratively add the element that minimizes f for each step up to the size of $\alpha k \ln(f(S_0)/E)$ (see Thm 1)
- To show the bounds on supermodular minimization, there is some construction toward Sparse Regression and Column Subset Selection where the intermediate results build toward the claim set out in the abstract.

Does the paper prove its claims?

- It does walk through the construction of

Exact setup of analysis/experiments

- Evaluates Clustering as a supermodular problem as introduction of methodology of general analysis
- Builds further evidence for its results via SMLR, Sparse Regression, and Column Subset Selection (relying heavily on literature).
- Key result is in improvement of Natarajan

Are there any gaps in the logic/proof?

- Jumps between full clustering objective being supermod. to constrained objective being supermod. (Lemma 2)
- No proof for Lemma 6

Possible next steps

- Determine exact or strong supermodularity expressions for a general (or as general as we can go) set function
- Applying results of this paper empirically.

2 Singer, Balkanski, et. al. 2016

What is the problem?

- Develops a new model of two-stage submodular maximization, learning sparse combinatorial representations.
- Multi-objective summarization

Why is it important?

- Provides a richer, more general, framework for submodular maximization
- Allows one to solve structured/hierarchical problems where the underlying functions are submodular themselves, a complication that causes the two-stage objective to be no longer submodular.

Why is it hard?

- If underlying functions are submodular, the approximation guarantees of the greedy algorithms no longer hold when subset selection (choosing S such that $|S| < l$) occurs at a higher stage.

Why existing solutions do not work?

- See above

What is the core intuition for the solution?

- Continuous relaxation of two-stage problem, being able to apply continuous greedy algorithm

Solution step-by-step?

- Two approaches, continuous relaxation where fractional solutions are interpreted as correlated distributions AND (when k is small) local search which initializes a solution suboptimally swapping elements to iteratively improve a potential function
- Continuous greedy algorithm provides fractional solution, mitigated by novel randomized rounding method

Does the paper prove its claims?

- Lemma 3.1 justifies the use of the continuous greedy algorithm
- Glosses over dependent rounding...

Exact setup of analysis/experiments

- Two experiments of both methods, one on a set of images, the other on a set of Wikipedia pages.
- Functions f_i are coverage functions for the Wikipedia data set, more general functions for the image data set

Are there any gaps in the logic/proof?

- Not that I could see, the paper is pretty complete and self contained (much of the logic underlying the intuition is moved to the appendix)

Possible next steps

- In view of the restricted Indian Buffet Process... What if we let k differ accross each underlying function f_j?