CSC2420: Algorithm Design, Analysis and Theory Spring 2019

Allan Borodin

January 17, 2019

Week 2

Announcements:

- There is one question so far on Assignment 1 and later today I will add a second question (or maybe more).
- Can we reschedule next weeks lecture sometime M, T, W??

Todays agenda

- The priority model as a model for greedy/myopic algorithms
- Some inapproximations with respect to priority algorithms; weighted interval selection.
- Greedy algorithms for the set packing problem and the s-set packing problem.
- $k+1$ claw-free graphs.
- The natural vertex cover and set cover greedy algorithms.
- Priority algorithms with revocations.
- Chordal graphs; perfect elimination ordering.
- Priority stack algorithms.

The priority algorithm model and variants

As part of our discussion of greedy (and greedy-like) algorithms, I want to present the priority algorithm model and how it can be extended in (conceptually) simple ways to go beyond the power of the priority model.

- What is the intuitive nature of a greedy algorithm as exemplified by the CSC 373 algorithms we mentioned? With the exception of Huffman coding (which we can also deal with), like online algorithms, all these algorithms consider one input item in each iteration and make an irrevocable "greedy" decision about that item..
- We are then already assuming that the class of search/optimization problems we are dealing with can be viewed as making a decision D_k about each input item I_k (e.g. on what machine to schedule job I_k in the makespan case) such that $\{(I_1, D_1), \ldots, (I_n, D_n)\}$ constitutes a feasible solution.
- For online problems (where the adversary determines the ordering of input item), the abstract problem formulation is called *request-answer games*. Note: The *line-search problem* and other online navigational search problems are not request-answer games.

Priority model continued

- Note: that a problem is only fully specified when we say how input items are represented. (This is usually implicit in an online algorithm.)
- We mentioned that a "non-greedy" online algorithm for identical machine makespan can improve the competitive ratio; that is, the algorithm does not always place a job on the (or a) least loaded machine (i.e. does not make a greedy or locally optimal decision in each iteration). It isn't always obvious if or how to define a "greedy" decision but for many problems the definition of greedy can be informally phrased as "live for today" (i.e. assume the current input item could be the last item) so that the decision should be an optimal decision given the current state of the computation.

Greedy decisions and priority algorithms continued

- For example, in the knapsack problem, a greedy decision always takes an input if it fits within the knapsack constraint and in the makespan problem, a greedy decision always schedules a job on some machine so as to minimize the increase in the makespan. (This is somewhat more general than saying it must place the item on the least loaded machine.)
- If we do not insist on greediness, then priority algorithms would best have been called myopic algorithms.
- We have both fixed order priority algorithms (e.g. unweighted interval scheduling and LPT makespan) and adaptive order priority algorithms (e.g. the set cover greedy algorithm and Prim's MST algorithm).
- The key concept is to indicate how the algorithm chooses the order in which input items are considered. We cannot allow the algorithm to choose say "an optimal ordering".
- We might be tempted to say that the ordering has to be determined in polynomial time but that gets us into the "tarpit" of trying to prove what can and can't be done in (say) polynomial time.

The priority model definition

- We take an information theoretic viewpoint in defining the orderings we allow.
- Lets first consider deterministic fixed order priority algorithms. Since I am using this framework mainly to argue negative results (e.g. a priority algorithm for the given problem cannot achieve a stated approximation ratio), we will view the semantics of the model as a game between the algorithm and an adversary.
- Initially there is some (possibly infinite) set $\mathcal{J}$ of potential inputs. The algorithm chooses a total ordering π on $\mathcal{J}$. Then the adversary selects a subset $\mathcal{I} \subset \mathcal{J}$ of actual inputs so that $\mathcal{I}$ becomes the input to the priority algorithm. The input items $I_1, \ldots, I_n$ are ordered according to π.
- In iteration k for $1 \leq k \leq n$, the algorithm considers input item I_k and based on this input and all previous inputs and decisions (i.e. based on the current state of the computation) the algorithm makes an irrevocable decision D_k about this input item.

The fixed (order) priority algorithm template

$\mathcal{J}$ is the set of all possible input items
Decide on a total ordering π of $\mathcal{J}$
Let $\mathcal{I} \subset \mathcal{J}$ be the input instance
$S := \varnothing$ % S is the set of items already seen
$i := 0$ % $i = |S|$
while $\mathcal{I} \setminus S \neq \varnothing$ **do**
    $i := i + 1$
    $\mathcal{I} := \mathcal{I} \setminus S$
    $I_i := \min_\pi \{I \in \mathcal{I}\}$
    make an irrevocable decision D_i concerning I_i
    $S := S \cup \{I_i\}$
end

Figure: The template for a fixed priority algorithm

Some comments on the priority model

- A special (but usual) case is that π is determined by a function $f : \mathcal{J} \to \Re$ and and then ordering the set of actual input items by increasing (or decreasing) values $f()$. (We can break ties by say using the input identifier of the item to provide a total ordering of the input set.) N.B. We make no assumption on the complexity or even the computability of the ordering π or function f.
- NOTE: Online algorithms are fixed order priority algorithms where the ordering is given *adversarially*; that is, the items are ordered by the input identifier of the item.
- As stated we do not give the algorithm any additional information other than what it can learn as it gradually sees the input sequence.
- However, we can allow priority algorithms to be given some (hopefully easily computed) global information such as the number of input items, or say in the case of the makespan problem the minimum and/or maximium processing time (load) of any input item. (Some inapproximation results can be easily modified to allow such global information.)

The adaptive priority model template

$\mathcal{J}$ is the set of all possible input items
$\mathcal{I}$ is the input instance
$S := \varnothing$ % S is the set of items already considered
$i := 0$ % $i = |S|$
while $\mathcal{I} \setminus S \neq \varnothing$ **do**
    $i := i + 1$
    decide on a total ordering π_i of $\mathcal{J}$
    $\mathcal{I} := \mathcal{I} \setminus S$
    $I_i := \min_{\leq_{\pi_i}} \{I \in \mathcal{I}\}$
    make an irrevocable decision D_i concerning I_i
    $S := S \cup \{I_i\}$
    $\mathcal{J} := \mathcal{J} \setminus \{I : I \leq_{\pi_i} I_i\}$
    % some items cannot be in input set
end

Figure: The template for an adaptive priority algorithm

Inapproximations with respect to the priority model

Once we have a precise model, we can then argue that certain approximation bounds are not possible within this model. Such inapproximation results have been established with respect to priority algorithms for a number of problems but for some problems much better approximations can be established using extensions of the model.

1. For the weighted interval selection (a *packing problem*) with arbitrary weighted values (resp. for proportional weights $v_j = |f_j - s_j|$), no priority algorithm can achieve a constant approximation (respectively, better than a 3-approximation).
2. For the knapsack problem, no priority algorithm can achieve a constant approximation. We note that the maximum of two greedy algorithms (sort by value, sort by value/size) is a 2-approximation.
3. For the set cover problem, the natural greedy algorithm is essentially the best priority algorithm.
4. As previously mentioned, for deterministic fixed order priority algorithms, there is an $\Omega(\log m / \log \log m)$ inapproximation bound for the makespan problem in the restricted machines model.

More on provable limitations of the priority model

The above mentioned inapproximations are with respect to deterministic priority algorithms. For an adaptive algorithm, the game between an algorithm and an adversary can conceptually be naturally viewed an alternating sequence of actions;

- The adversary eliminates some possible input items
- The algorithm makes a decision for the item with highest priority and chooses a new ordering for all possible remaining input items.

However, we note that for deterministic algorithms, since the adversary knows precisely what the algorithm will do in each iteation, it could initially set the input $\mathcal{I}$ once the algorithm is known.

More on provable limitations of the priority model

The above mentioned inapproximations are with respect to deterministic priority algorithms. For an adaptive algorithm, the game between an algorithm and an adversary can conceptually be naturally viewed an alternating sequence of actions;

- The adversary eliminates some possible input items
- The algorithm makes a decision for the item with highest priority and chooses a new ordering for all possible remaining input items.

However, we note that for deterministic algorithms, since the adversary knows precisely what the algorithm will do in each iteation, it could initially set the input $\mathcal{I}$ once the algorithm is known.

For randomized algorithms, there is a difference between an *oblivious adversary* that creates an initial subset $\mathcal{I}$ of items vs an *adaptive adversary* that is playing the game adaptively reacting to each decision by the algorithm.

Unless stated otherwise we usually analyze randomized algorithms (for any type of algorithm) with respect to an oblivious adversary.

Interval selection and greedy priority algorithms

To illustrate the limitations of greedy algorithms (as modeled by priority algorithms), we consider the (weighted) interval selection problem. Here we are given n intervals $I_1, \ldots, I_n$ where intervals are given by $I_j = [s_j, f_j)$. I_j and I_k are conflicting if $I_j \cap I_k \neq \varnothing$. (I use half closed, half open intervals to indicate that we allow intervals to intersect just at an endpoint.) Each interval I_j has a weight or value v_j. The goal is to find a subset of non-conflicting intervals so as to maximize the sum of the values of the selected intervals.

It is well known that for the unweighted case (i.e., when $v_j = 1$ for all j) that the fixed order greedy priority algorithm Earliest Finishing Time (EFT) is optimal.

Priority algorithms for one machine interval selection continued

1. For proportional values (i.e., when $v_j = f_j - s_j$ for all j), the Longest Processing Time (LPT) greedy algorithm achieves a 3-approximation which can be proven by a charging argument. This is essentially the best possible for any deterministic priority algorithm as shown onthe following slide.
   Note: Perhaps surprinsingly for $m = 2$ machines, there is a priority 2-approximation algorithm. This algorithm can be re-stated to provide a randomized priority 2-approximation.
2. For arbitrary values $\{v_j\}$, no deterministic priority algorithm can achieve a constant approximation. More precisely, if we let $\delta_j = \frac{v_j}{f_j - s_j}$ and $\Delta = \frac{\max_j \delta_j}{\min_j \delta_j}$, then no deterministic priority algorithm can achieve ratio better than Δ. The LPT algorithm yield a 3Δ approximation for arbitrary profits.
   Note: If $\min_i \delta_i$ and $\max_i \delta_i$ are known then there is a randomized priority $O(\log \Delta)$ approximation.

A simple priority algorithm inapproximation for proportional profit

The nemisis sequence consists of long and short jobs. The long jobs are depicted in the figure. There is a small ϵ overlap between intervals on the top and bottom. In addition for each long interval of length (= value) v_i, there are three short intervals of length $\frac{v_i - 2\epsilon}{3}$ included in that long interval. The adversary will be able to create a subset of these intervals to force a bound arbitrarily close to 3 (for sufficiently small ϵ and large q).

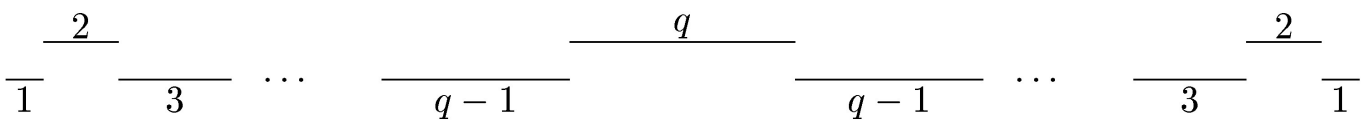

Figure: The long jobs in the nemesis input

The first interval (which must be selected) considered by the priority algorithm will allow the adversary to remove enough items to force the bound. There are four cases: I_1 is a short interval, I_1 has length 1, I_1 has length ℓ for $1 < \ell < q$, and I_1 has length q.

Extensions of the priority order model

In discussing more general online frameworks, we already implicitly suggested some extensions of the basic priority model (that is, the basic model where we have one-pass and one irrevocable decision). The following online or priority algorithm extensions can be made precise:

- Decisions can be *revocable* to some limited extent or at some cost. For example, we know that in the basic priority model we cannot achieve a constant approximation for weighted interval scheduling. However, if we are allowed to permanently discard previously accepted intervals (while always maintaining a feasible solution), then we can achieve a 4-approximation. (but provably not optimality).
- While the knapsack problem cannot be approximated to within any constant, we can achieve a 2-approximation by taking the maximum of 2 greedy algorithms. More generally we can consider some "small" number k of priority (or online) algorithms and take the best result amongst these k algorithms. The partial enumeration greedy algorithm for the makespan and knapsack problems are an example of this type of extension.

Extensions of the priority order model continued

- Closely related to the "best of k online" model is the concept of online algoitthms with "advice". (One could also study priority algorithms with advice but that has not been studied extensively.) There are two advice models, a model where one measures the maximum number of advice bits per input item, and a model where we are given some number ℓ of advice bits at the start of the computation. The latter model is what I will mean by "online with advice." Online with ℓ advice bits is equivalent to the max of $k = 2^\ell$ online model.

Extensions of the priority order model continued

- Closely related to the "best of k online" model is the concept of online algoitthms with "advice". (One could also study priority algorithms with advice but that has not been studied extensively.) There are two advice models, a model where one measures the maximum number of advice bits per input item, and a model where we are given some number ℓ of advice bits at the start of the computation. The latter model is what I will mean by "online with advice." Online with ℓ advice bits is equivalent to the max of $k = 2^\ell$ online model.
  NOTE: This model is a very permissive in that the advice bits can be a function of the entire input. Of course, in practice we want these advice bits to be "easily determined" (e.g., the number of input items, or the ratio of the largest to smallest weight/value) but in keeping with the information theoretic perspective of onine and priority algorithms, one doesn't impose any such restriction.
- There are more general parallel priority based models than "best of k" algorithms. Namely, parallel algorithms could be spawning or aborting threads (as in the pBT model to be discussed later).

Multipass algorithms

- Another model that provides improved results is to allow multiple passes (over the input items) rather than just one pass.
- This is not a well studied model but there are two relatively new noteworthy results that we will be discussing:
  1. There is deterministic 3/4 approximation for weighted Max-Sat that is achieved by two "online passes" (i.e., the input sequence is determined by an adversary) over the input sequence whereas there is evidence that no one pass deterministic online or priority algorithm can acheive this ratio.
  2. There is a $\frac{3}{5}$ approximation for biparitie matching that is achieved by two online passes whereas no deterministic online or priority algorithm can do asymptotically better than a $\frac{1}{2}$ approximation.
- It is not clear how best to formalize these multi-pass algorithms. Why?

Multipass algorithms

- Another model that provides improved results is to allow multiple passes (over the input items) rather than just one pass.
- This is not a well studied model but there are two relatively new noteworthy results that we will be discussing:
  1. There is deterministic 3/4 approximation for weighted Max-Sat that is achieved by two "online passes" (i.e., the input sequence is determined by an adversary) over the input sequence whereas there is evidence that no one pass deterministic online or priority algorithm can acheive this ratio.
  2. There is a $\frac{3}{5}$ approximation for biparitie matching that is achieved by two online passes whereas no deterministic online or priority algorithm can do asymptotically better than a $\frac{1}{2}$ approximation.
- It is not clear how best to formalize these multi-pass algorithms. Why? What information should we be allowed to convey between passes?

Greedy algorithms for the set packing problem

One of the new areas in theoretical computer science is algorithmic game theory and mechanism design and, in particular, auctions including what are known as *combinatorial auctions*. The underlying combinatorial problem in such auctions is the set packing problem.

The set packing problem

We are given n subsets $S_1, \ldots, S_n$ from a universe U of size m. In the weighted case, each subset S_i has a weight w_i. The goal is to choose a disjoint subcollection $\mathcal{S}$ of the subsets so as to maximize $\sum_{S_i \in \mathcal{S}} w_i$. In the s-set packing problem we have $|S_i| \leq s$ for all i.

- This is a well studied problem and by reduction from the max clique problem, there is an $m^{\frac{1}{2}-\epsilon}$ hardness of approximation assuming $NP \neq ZPP$. For s-set packing with constant $s \geq 3$, there is an $\Omega(s/\log s)$ hardness of approximation assuming $P \neq NP$.
- We will consider two "natural" greedy algorithms for the s-set packing problem and a non obvious greedy algorithm for the set packing problem. These greedy algorithms are all fixed order priority

The first natural greedy algorithm for set packing

Greedy-by-weight ($Greedy_{wt}$)

Sort the sets so that $w_1 \geq w_2 \ldots \geq w_n$.
$\mathcal{S} := \varnothing$
For $i : 1 \ldots n$
    If S_I does not intersect any set in $\mathcal{S}$ then
        $\mathcal{S} := \mathcal{S} \cup S_i$.
End For

- In the unweighted case (i.e. $\forall i, w_i = 1$), this is an online algorithm.
- In the weighted (and hence also unweighted) case, greedy-by-weight provides an s-approximation for the s-set packing problem.
- The approximation bound can be shown by a charging argument where the weight of every set in an optimal solution is charged to the first set in the greedy solution with which it intersects.

The second natural greedy algorithm for set packing

Greedy-by-weight-per-size

Sort the sets so that $w_1/|S_1| \geq w_2/|S_2| \ldots \geq w_n/|S_n|$.
$\mathcal{S} := \varnothing$
For $i : 1 \ldots n$
    If S_I does not intersect any set in $\mathcal{S}$ then
        $\mathcal{S} := \mathcal{S} \cup S_i$.
End For

- In the weighted case, greedy-by-weight provides an s-approximation for the s-set packing problem.
- For both greedy algorithms, the approximation ratio is tight; that is, there are examples where this is essentially the approximation. In particular, these algorithms only provide an m-approximation where $m = |U|$.
- We usually assume $n >> m$ and note that by just selecting the set of largest weight, we obtain an n-approximation. So the goal is to do better than $\min\{m, n\}$.

Improving the approximation for set packing

- In the unweighted case, greedy-by-weight-per-size can be restated as sorting so that $|S_1| \leq |S_2| \ldots \leq |S_n|$ and it can be shown to provide an $\sqrt{m}$-approximation for set packing.
- On the other hand, greedy-by-weight-per-size does not improve the m-approximation for weighted set packing.

Greedy-by-weight-per-squareroot-size

Sort the sets so that $w_1/\sqrt{|S_1|} \geq w_2/\sqrt{|S_2|} \ldots \geq w_n/\sqrt{|S_n|}$.
$\mathcal{S} := \varnothing$
For $i : 1 \ldots n$
    If S_I does not intersect any set in $\mathcal{S}$ then
        $\mathcal{S} := \mathcal{S} \cup S_i$.
End For

Theorem: Greedy-by-weight-per-squareroot-size provides a $2\sqrt{m}$-approximation for the set packing problem. And as noted earlier, this is asymptotically the best possible approximation assuming $NP \neq ZPP$.