Analysis of Pipage method for k-Max Coverage

work with Prof. Robert Kleinberg and Lisa Fawcett

January 16, 2012

1 Introduction

Fractional k–max coverage has a polynomial time algorithm due to linear programing. It is natural and tempting to convert a fractional solution to an integral one by a series of exchanges that do not decrease the objective. Ageev and Sviridenko formalized this method as *pipage* rounding. This method has a tight approximation ratio of $1 - \frac{1}{e}$ for k-max coverage. However, in most examples, in seem to perform much better. Our goal is to answer questions such as: Can we explain why *pipage* is much better on average? Can we limit their worst case analysis to a special class of inputs, such as set systems with high VC-dimension? Is there a more efficient variant of pipage moves with ε-competitiveness?

2 Some proofs

Let's characterize what the "worst case" input for Pipage looks like. By worst case, we mean that the optimal solution x^* to L-objective satisfies

$$F(x^*) \leq (C + \varepsilon)L(x^*)$$

for some constant $\varepsilon > 0$ (the competitiveness factor that we are aiming for). Let's call an input with such property an ε-worst input.

To more precisely state what we want, we need to introduce a bit of notation. We use $\forall^\varepsilon$ to denote "all but" $\delta(\varepsilon)$ fraction, where $\delta(\varepsilon) \to 0$ as $\varepsilon \to 0$. Similarly, $x =^\varepsilon y$ means $|x - y| < \delta(\varepsilon)y$. One exception is $x =^\varepsilon 0$, which means instead $|x| < \delta(\varepsilon)$ (this is for convenience, and is still reasonable).

Theorem 2.1. *An ε-worst input must have $\forall^\varepsilon$ elements i, either:*

1. $\sum_{S \ni i} x^*_S =^\varepsilon 0$, *or*
2. $\sum_{S \ni i} x^*_S =^\varepsilon 1$ *(ie. fully covered)*

*Also, define **good** elements to be all elements that satisfy condition 2*

Proof. This is from the last step of deriving constant C in $F(x, ..., x_n) \geq CL(x_1, ..., x_n)$, where they showed that for an element j, the associated function $g(z)$ satisfies

$$g(z) = 1 - \left(1 - \tfrac{z}{p}\right)^p \geq (1 - (1 - 1/p)^p)z$$

(where $z = \min\{1, \sum_{i=1}^{p} y_i\}$ and y_i or renamed x_i of the subsets that cover element j) since $g(z) = 1 - (1 - z/p)^p$ is concave on the segment $[0, 1]$.

The multiplicative error $e(z) = \dfrac{1 - (1 - z/p)^p}{(1 - (1 - 1/p)^p)z} = O(1 - z)$ when we're measuring the error with respect to $z = 1$ since $e(z)$ is a polynomial of degree $p - 1$ but the first-order term dominates since $0 \le z \le 1$.
The additive error $e(z) = (1 - (1 - z/p)^p) - (1 - (1 - 1/p)^p)z = O(z)$ similarly with respect to $z = 0$.

Using our new notation, with $\delta_1(\varepsilon) = \varepsilon^{1/2}$ and $\delta_2(\varepsilon) = \varepsilon^{1/2}$, we see that for all but $\delta_1(\varepsilon)$ fraction of elements $z =^{\varepsilon} 0$ or $z =^{\varepsilon} 1$ (with respect to δ_2).

If $z = \min\{1, \sum_{i=1}^{p} y_i\} =^{\varepsilon} \sum_{i=1}^{p} y_i$, we have $\sum_{i=1}^{p} y_i =^{\varepsilon} 1$ by transitivity.
Otherwise, recall that in their use of AM-GM inequality to show

$$1 - \prod_{i=1}^{p} (1 - y_i) \ge (1 - (1 - 1/p)^p) \sum_{i=1}^{p} y_i \ge (1 - (1 - 1/p)^p) \min\{1, \sum_{i=1}^{p} y_i\}$$

there was already enough gap between z and $\sum_{i=1}^{p} y_i$ □

Lemma 2.2. *(Multiplicative error of AM-GM) Given n numbers x_i in $[0, 1]$. If*

$$\frac{\prod x_i}{\left(\dfrac{\sum x_i}{n}\right)^n} \ge (1 - \varepsilon)$$

then $\forall^{\varepsilon}$ $x_i =^{\varepsilon} \overline{x}$, where $\overline{x} = \sum x_i / n$

Proof. Consider as if we start initially with the numbers $\overline{x}, ..., \overline{x}$, and transform it to $x_1, ..., x_n$ by a series of exchanges between two terms. In an exchange, we pick $x_i \ge \overline{x}$ and $x_j \le \overline{x}$ (note, $x_i \ge x_j$), and adjust it to $x_i = x_i + d$ and $x_j = x_j - d$ where $d \le \min(1 - x_i, x_j)$. It is easy to see that such sequence of exchanges always exists that allow us to reach the final x_i's.

Initially, we start with $\prod x_i = \dfrac{\sum x_i}{n}^n$. At each exchange, the overall product decreases multiplicatively by:

$$\frac{(x_i + d)(x_j - d)}{x_i x_j} \le 1 - \frac{d^2}{x_i x_j} \le 1 - d^2$$

Choice of $\delta_1(\varepsilon) = n\varepsilon^{1/4}$ and $\delta_2(\varepsilon) = 2\varepsilon^{1/4}$ finishes the proof. To see this work, suppose $\delta_1(\varepsilon)$ fraction of numbers are off by more than $\delta_2(\varepsilon)$, for a total offset of $2n\varepsilon^{1/2}$, then there is at least one number $x_i \ge \overline{x} + \varepsilon^{1/2}$ and one number $x_j \le \overline{x} - \varepsilon^{1/2}$ by pigeonhole. With an exchange move on these two numbers with $d = \varepsilon^{1/2}$, the overall product decreases by at least $1 - d^2 = 1 - \varepsilon$, a contradiction. □

Theorem 2.3. *An ε-worst input must have $\forall^{\varepsilon}$ good elements i:*

$$\# \text{ sets containing } i = \frac{n}{k}$$

and moreover,

$$\forall^{\varepsilon} S \ni i \;\; x_s^* =^{\varepsilon} \frac{k}{n}$$

Proof. (as a reminder, n is the total number of subsets, and k the number of subsets we are allowed to pick).

For each good element j that we are considering, with respect to the p subsets that cover the element, we have

$$1 - \prod_{S_i \ni j} (1 - x_i) =^{\varepsilon} (1 - (1 - 1/p)^p)(\sum_{i=1}^{p} x_i)$$

Because we are considering only the good elements, the sum is approximately 1 by Theorem 2.1. From this and Lemma 2.2, we deduce that $x_i =^{\varepsilon} 1/p$. However, if a subset i covers another element j', then by the same argument we have $x_i =^{\varepsilon} 1/p'$, where p' is the number of subsets that cover element j'.
Since there is only one value of x_i, we must have $p =^{\varepsilon} p'$.

We continue with this reasoning for the rest of the set system. It is easiest to describe in terms of graph, so we translate the set system to a multi graph, where each node is a subset, and two nodes are connected by an edge if their corresponding subsets share an element.

From above, we have that all subsets in the same "connected component" of this graph must have ε–equal values of x_i, that is, all the subsets are picked approximately equally.

It suffices to assume there is only one component, since if we have a poly-time algorithm for 1-component set system (a k-component set system is one whose graph has k connected components), we can extend it to k-components easily by first computing k'-max coverage for each component separately with varying values of k', and running the KNAPSACK algorithm. Since each of the components are independent, each (k', component) pair is an item with cost k' and payoff equal to the value of k'-max coverage on the component. The bound on the total knapsack size is k. (the knapsack algorithm will have polynomial runtime, unlike in the general case, since the weights of the items are bounded polynomially by the size of the set system).

Finally, since $\sum x_i = k$ and they are ε-equal, $x_i =^{\varepsilon} \frac{k}{n}$. Then, we also know that # sets containing any good element is $=^{\varepsilon} \frac{n}{k}$. □

3 Goal

Let g be the number of good elements. We want to show that unless $\forall^{\varepsilon} S \; |S| =^{\varepsilon} \frac{g}{k}$, pipage rounding succeeds (or at least there exists a sequence of pipage moves) in improving F by $1 + c$ for some constant c.

Suppose $|S_i| > (1 + \delta)|S_j|$ (or $(1 - \delta)|S_i| > |S_j|$). Look at ΔF as we change x^* to $x^* + \Delta x$ where:

$$\begin{cases} +\gamma & \text{if } s_i = s \\ -\gamma & \text{if } s_j = s \\ 0 & \text{if otherwise} \end{cases}$$

We want to analyze how this infinitesimal change in x affects F. We do case analysis on element e:

- $e \in S_i \cup S_j$. Contribution to $F(x^*)$ was $1 - (1 - \frac{k}{n})^{\frac{n}{k}}$, is now $1 - (1 - \frac{k}{n})^{\frac{n}{k}} \cdot \dfrac{(1-\frac{k}{n})^2\,\gamma^2}{(1-\frac{k}{n})^2}$
- $e \in S_i \setminus S_j$. Contribution to $F(x^*)$ was $1 - (1 - \frac{k}{n})^{\frac{n}{k}}$, is now $1 - (1 - \frac{k}{n})^{\frac{n}{k}} \cdot \dfrac{1-\frac{k}{n}-\gamma^2}{1-\frac{k}{n}}$
- $e \in S_j \setminus S_i$. Contribution to $F(x^*)$ was $1 - (1 - \frac{k}{n})^{\frac{n}{k}}$, is now $1 - (1 - \frac{k}{n})^{\frac{n}{k}} \cdot \dfrac{1-\frac{k}{n}+\gamma^2}{1-\frac{k}{n}}$

<u>Sketch of what we might try</u>
For one pipage operation the benefit is $c\delta|S_i|\frac{k}{n}$ for some appropriate constant c. After n operations, total benefit is $c \cdot \delta\frac{k}{n} \cdot$ (total size of sets used). Since every $|S_i| \geq \frac{g}{k}$, this is at least $c\delta g$. Of course, the gradient was evaluated only at the initial starting point, so our analysis is only a rough estimate. (this is the one section I'm still slightly unsure where it's leading to)

3.1 Plan: towards a characterization of the sets

Why might it be unlikely for an ε-worst set system to exist? Can such property imply *high* VC-dimension, or some non-trivial structural property?

Our long range plan is to show that integrality gap smaller than $(1 - 1/e)(1 + \varepsilon)$ implies that VC-dimension $d > f(\varepsilon)$ for some f that satisfies $f \to \infty$ as $\varepsilon \to 0$.

Lemma 3.1. *If you have 2^d sets and their Venn diagram is full, then they shatter a set of d elements*

Proof. We can represent the set system as a $2^{2^d} \times 2^d$ matrix, where each column represents a set, and each row represents an element; $a_{ij} = 1$ if set j contains the element i. Since the Venn diagram is full we have at least 2^{2^d} unique elements, each of which is represented by a unique binary vector over the sets.

If we look at the following subset of rows (example $d = 3$),

$$\begin{pmatrix} 0 & 0 & 0 & 0 & 1 & 1 & 1 & 1 \\ 0 & 0 & 1 & 1 & 0 & 0 & 1 & 1 \\ 0 & 1 & 0 & 1 & 0 & 1 & 0 & 1 \end{pmatrix}$$

we can easily verify that the d elements corresponding to those rows are shattered, since there is a subset containing every combination. A similar construction can be used for the general case. □

3.2 A Parallel plan

Because the SET COVER problem has the dual formulation of HITTING SET (swap the roles of sets and elements), we can also sketch out the following plan of attack.

For any t–tuple of elements where $t \leq f(\varepsilon)$, the "dual Venn diagram" of those t elements looks like t independent random sub–collections of the sets, each containing $\frac{1}{k}$ fraction. An advantage of this method is that there is no log-factor blow up.

References

1. Alexander A. Ageev and Maxim I. Sviridenko. Approximation Algorithms for Maximum Coverage and Max Cut with Given Sizes of Parts. IPCO'99, LNCS 1610, pp. 17-30, 1999.