Regret calculations

@benthestatistician

May 20, 2021

Problem formulation

The given problem is to find a flow

$$\vec{x}^* = \{x^*_{ij} : (i, j) \in \mathcal{A}\}$$

that, among flows $\vec{x} \in \mathcal{C}$ satisfying constraints, minimizes

$$\sum_{(i,j)\in\mathcal{A}} a_{ij} x_{ij}. \tag{1}$$

The arc set $\mathcal{A}$ often contains a subset $\mathcal{B}$ of bookkeeping arcs that do not themselves indicate the presence or absence of a match. In the standard representation of full matching problems, $a_{ij} = 0$ for all $(i, j) \in \mathcal{B}$, making (1) the same as $\sum_{(i,j)\in\mathcal{A}\setminus\mathcal{B}} a_{ij} x_{ij}$.

Our method of solving this problem begins by up-shifting and discretizing arc costs:

$$\begin{aligned} \text{given } d_{ij} &:= \begin{cases} \epsilon\lceil 0.5 + a_{ij}/\epsilon\rceil, & (i,j) \in \mathcal{A}\setminus\mathcal{B} \\ 0, & (i,j) \in \mathcal{B} \end{cases} \\ \text{minimize} &\sum_{(i,j)\in\mathcal{A}} d_{ij} x_{ij} \text{ over } \vec{x} \in \mathcal{C} \end{aligned} \tag{2}$$

for some $\epsilon > 0$. The up-shift ensures all costs are positive, while the discretization satisfies a requirement of the solver. Given that $\vec{x}^*$ minimizes (2), what can we say about

$$\sum_{(i,j)\in\mathcal{A}\setminus\mathcal{B}} a_{ij} x^*_{ij} - \min_{\vec{x}\in\mathcal{C}} \sum_{(i,j)\in\mathcal{A}\setminus\mathcal{B}} a_{ij} x_{ij}, \tag{3}$$

the *regret*? Here are two general strategies.

Minorization

Let f be an real-valued function to be minimized over $\mathcal{S}$. Let g be another function on the same domain. Say g minorizes f if $g \leq f$. By minimizing g we don't necessarily minimize f, but we obtain a lower bound for the infimum of f. Specifically, if $x^* = \arg\min_{x \in \mathcal{S}} g(x)$, then $f(x^*) - \min_{x \in \mathcal{S}} f(x) \leq f(x^*) - g(x^*)$.

I claim that

$$\sum_{(i,j) \in \mathcal{A} \setminus \mathcal{B}} (d_{ij} - 1.5\epsilon) x_{ij} \text{ minorizes } \sum_{(i,j) \in \mathcal{A} \setminus \mathcal{B}} d_{ij} x_{ij},$$

because $d_{ij} - 1.5\epsilon \leq a_{ij}$. (It can be checked that $x \mapsto x - 1.5\epsilon$ is the smallest downward shift enabling minorization.)

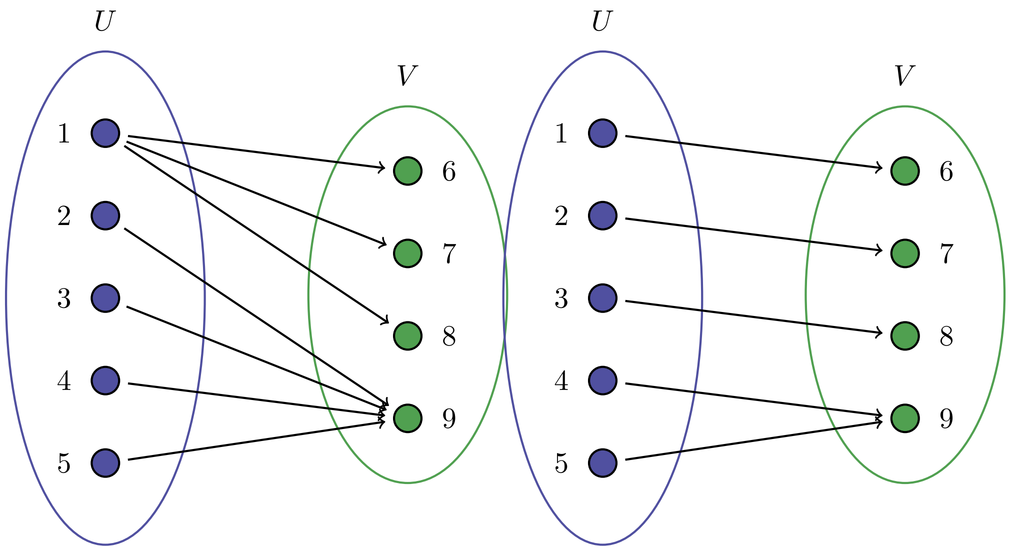
Does minimization of (2) entail minimization of $\sum_{(i,j) \in \mathcal{A} \setminus \mathcal{B}} (d_{ij} - 1.5\epsilon) x_{ij}$? A sufficient condition for this is that the minimizer $\vec{x}^*$ of (2) maximizes $\vec{x} \mapsto \sum_{(i,j) \in \mathcal{A} \setminus \mathcal{B}} x_{ij}$ over $\mathcal{C}$. In these cases we can find $\vec{x}^*$ minimizing (2), infer that $\vec{x}^*$ also minimizes $\sum_{(i,j) \in \mathcal{A} \setminus \mathcal{B}} (d_{ij} - 1.5\epsilon) x_{ij}$, then compare the values of (1) and this minimizing function at $\vec{x}^*$ for a regret bound.

Problems $(\mathcal{A}, \mathcal{B}, \mathcal{C})$ representing pair matching and matching with k controls enjoy the property that solutions of (2) maximize $\vec{x} \mapsto \sum_{(i,j) \in \mathcal{A} \setminus \mathcal{B}} x_{ij}$ over $\mathcal{C}$, in virtue of that function begin constant over $\vec{x} \in \mathcal{C}$. These problems having a separate invocation (`pairmatch()`) in optmatch, it would be sensible for optmatch regret calculators, and `summary.optmatch()`, to treat them a bit separately also.

But first we should consider prospects for saying something about full-matching problems also. These certainly permit that $\vec{x} \mapsto \sum_{(i,j) \in \mathcal{A} \setminus \mathcal{B}} x_{ij}$, which we might term the *number of active arcs*, not be at its maximum over $\mathcal{C}$ as $\sum_{(i,j) \in \mathcal{A}} d_{ij} x_{ij}$ is minimized over $\mathcal{C}$; solutions keeping the number of active arcs well below its maximum are generally preferred. For example, the "stability increments" discussed in Hansen and Klopfer (2006) add to the optimization objective a penalty proportional to the number of active arcs. In ordinary full matching problems, more active arcs means more sharing of controls; see Figure 1. If we can conceptualize minimization of the sharing of controls as a secondary matching objective, then our minorization bound becomes interpretable as a bound on regret relative to a Pareto optimum of the original problem. (Pareto regret?)

This "Pareto regret" story applies to ordinary full matching problems, i.e. problems that:

1. may or may not set `omit.fraction` to a nonzero value, but if they do



$$\sum_{(i,j)\in\mathcal{A}\setminus\mathcal{B}} x_{ij} = 7;\ n_{\mathrm{eff}} = 3.1 \qquad \sum_{(i,j)\in\mathcal{A}\setminus\mathcal{B}} x_{ij} = 5;\ n_{\mathrm{eff}} = 4.3$$

Figure 1: Two candidate solutions to a hypothetical unrestricted full matching problem. If the candidate at left achieves a smaller sum of matched distances, it does so at the cost of greater reuse of controls, leading to its effective sample size, $n_{\mathrm{eff}} = \sum_s [(n_{st}^{-1} + n_{sc}^{-1})/2]^{-1}$, being smaller.

then they set it to a positive value, indicating that it's a fraction of the eligible control reservoir that's to be left out;

2. may or may not impose `max.controls` limits on treatment:control ratios by matched set, but if they do those limits do not require sharing of controls (`max.controls` $\leq 1/2$).

A (sub)-problem not meeting both of Conditions 1 and 2 is termed "flipped," and is handled by flipping the roles of the treatment and control groups before transforming the matching problem into its min-cost flow representation $(\mathcal{A}, \mathcal{B}, \mathcal{C})$. Consequently, in flipped subproblems larger numbers of active arcs correspond to larger numbers of control subjects sharing the same matched treatment-group counterpart.

This complicates our "Pareto regret" story at least somewhat. The secondary objective needs to become something along the lines of:

Minimize sharing of controls by multiple members of the treat-

> ment group — unless treatments are so plentiful relative to controls that some are being discarded, in which case it's the sharing of treatment group members by multiple members of the control group that's to be minimized.

(If the problem was flipped because it fails Condition 2, then its forbids one-many [one treatment/multiple controls] matched sets, and the secondary objective is automatically met. So the one edge case we're left worrying about is failure of Condition 1, i.e. negative omit.fractions.) Without flipping, one aims for one-many matched sets only, or failing that to keep many-one matched sets to a minimum; with flipping it's the reverse.

Can this secondary objective be posed in terms of effective sample size? Something along those lines making it simpler to state?

Duality

(Duality-based regret bounds flow from standard sources. Still, remains to tell the story.)

References

Hansen, B. B. and Klopfer, S. O. (2006), "Optimal full matching and related designs via network flows," *Journal of Computational and Graphical Statistics*, 15, 609–627.