Optimizing spatial survey for Chagas vectors

UPenn-UIC Chagas Team

1 Notation

- P, the set of houses in the geographic area
- $w(p)$, the risk associated with a location $p \in P$(=importance to know this spot)
- $v_i \in \{0, 1\}$, the indicator whether house $i \in P$ is selected for a surveillance visit
- b, the surveillance budget in houses per "arm pull" [Sasha: I think of this as the budget over the term of the optimization = the overall number of houses we would ever explore]
- $f_p(i) : P \to \Re^+$, the information of i about p
  Corentin: I inversed the indexes to correspond to the following is it right? [Sasha: yes, and good notation]
- $g(.)$, return of investment function, a concave increasing function, like $\frac{x}{1+x}$.

2 Model for exploratory review (PSU version)

$$
\begin{aligned}
v &= \{v_p\}_{p \in P} && (1) \\
&= \underset{v}{\operatorname{argmax}} \sum_{p \in P} w(p)\ g\left(\sum_{i \in P} f_p(i) v_i\right) && (2) \\
subject \quad to \quad & \sum_i v_i = b && (3)
\end{aligned}
$$

A slight generalization: g could take multiple arguments like $g_p(v_{p_1}, v_{p_2}, \dots)$, where $\{p_j\} = N_p$ are some points in the vacinity of p.

The model could be used in the following way.

1. it is solved to near optimality. Greedy algorithm should give probably good solution.
2. the survey team would systematically visit the houses in the surveillance solution ($v_i = 1$).

3. if there is no detection, the team proceeds to the next house.

4. If there is detection at site i, then the exploration is suspended, and the team proceeds to investigate K houses AROUND site i. The value of K is determined separately.

3 Model options for explotation

Set radius - say within 50 meters

alternatives

Multiple radiuses as different arms - within 20 m - within 20-40m - within 40-70m

-Exploitation could be adaptive go to immediate neighbors (a) and then secondary neighbors (b) until all in set b are negatives. first neighbor 13 meters ¡17 meters

1.

3.1 Finding T. cruzi

Search for human cases. Finding as many infected humans as possible.

- $w(p)$, potentially detections in humans, guinea pigs, dogs
- Select one of them based on the Guadalupe data and proof of principle either on the real data or on simulated data "alike". See: Figure 3 of this paper: http://www.ploscompbiol.org/article/infoignoring the triangles etc. Trick is to find the red dots (cases) PMC3174153

The better you are in exploitation, the more you should spend on it! (larger K).

```
X   X
 H  X
X   X
```

3.2 Finding cases

we have serology from 1000 people from 250 houses in a small community. There are clear spatially-focal microepidemics. Idea would be to come up with an optimal serach algrithm to find these cases with no, or very little prior info. We could also include prior info of risk factors. Downside is this dataset has been analyzed to death. Upside is its easier to publish on disease.

3.3 Finding the bugs

We could use spray data as 50.000 houses with participation 60%. Arround 20% of infestation with serious variations (localities with near nothing and localities at 30%.

Shape of the autocorrelation for infestation:

```
\
 \
  \
   \_ _ _ _ _ _
               \
                \
   70m          600m
```

In a first setting we should be using exploration within 70m arround hits. Exploration on a grid, with possibility of densifying if some positive.

The search would be based on reports and extending around reports vs. exploring new regions.

Trial dataset could be: - space: (a) 100x100 grid houses. each house is 10m x 10m, (b) actual map (Corentin/Karthik will send ASAP a .csv with x,y,positive) - budget of 1000 total visits then evaluate - start with a bunch of spots - simulate the dispersal with hops model fit on Jerusalen data for 2 years. - leave out reporting

Outcomes - validation: the algorithm lays out a grid-like search pattern of v_i

- through some tuning of b and K, we could achieve a reasonable tradeoff between ET (exploit) and ER (explore)

Then simulate infested people to report at a rate of 1/10 per year. And see what is the best strategy to controle that. (this is modeled as a weight $w(p)$)

Good bedtime reading[1]

References

[1] D. Adams. *The Hitchhiker's Guide to the Galaxy.* San Val, 1995.