AM207 Final Project Proposal: Modelling Civil Conflict and Refugee Relief Aid in Uganda

Peter J. Bull
Institute for Applied Computer Science
Harvard University
Cambridge, Massachusetts
bull@fas.harvard.edu

Isaac M. Slavitt
Institute for Applied Computational Science
Harvard University
Cambridge, Massachusetts
slavitt@fas.harvard.edu

Abstract

Using stochastic modeling techniques learned in AM207, we aim to model civil conflict in Uganda and optimize the limited relief aid that can be provided in these scenarios. The goal is two-fold: first, to simulate civil conflict events given historical data about these events. This could have the benfts of predict future incidents and modelling the spread of violence during larger-scale civil conflicts.

The second goal is to optimize the distribution of limited aid resources to refugees from these crises. Using features about the geography, demographics and infrastructure of Uganda, we can attempt to find optimal locations to deliver aid so that the greatest number of refugees from these crises can be served. We will approach this complex optimization task using stochastic methods from the course.

1   The Data

The data comes from ACLED (Armed Conflict Location and Event Data Project) [1], which is a dataset with locations, dates, fatalities, motivation, actors involved, and other information about civil conflicts in Africa. Their collection of data on Uganda covers 1997-2013, and they have a real-time tracker of events reported in 2014. This new data from 2014 could be used as a test set for our modeled simulations of violence. The need for an understanding of these patterns of conflict is clear, as ACLED notes:

> This dataset codes the dates and locations of all reported political violence events in over 50 developing countries. Political violence includes events that occur within civil wars or periods of instability. Although civil war occurrence is decreasing across African countries, new forms of political violence are becoming more common. [1]

Given that ACLED has collected data for civil conflict across many developing nations, I hope that the models built for this project can be tested, adapted, and applied to geographic areas outside of Uganda.

For Uganda, the dataset contains around 4500 observations of civil violence. Each observation includes the data, location, an quantified estimate for how precise these measures are, a number of fatalities for the event, and the actors involved.w
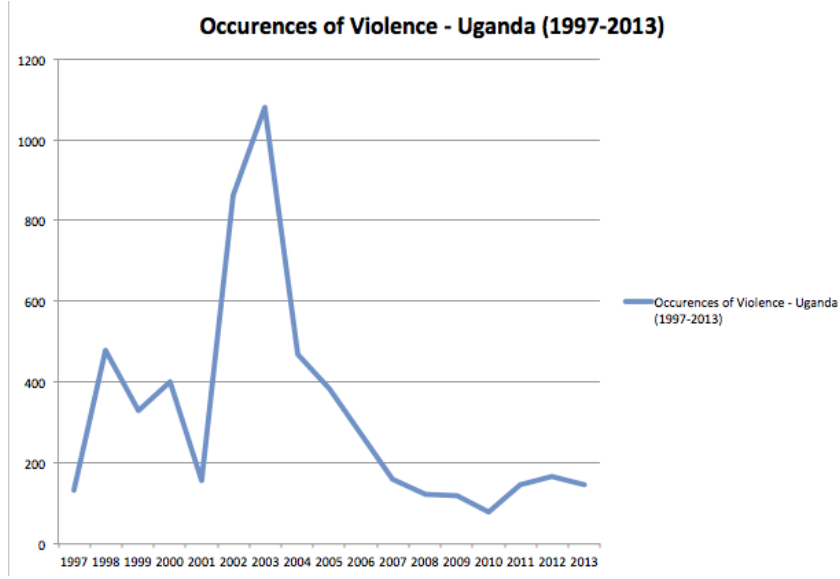
Figure 1: Civil conflict per year in Uganda 1997-2013. [2]

2 Modeling Civil Conflict Events

2.1 Bayesian Poisson Regression

We will use a bayesian monte-carlo model to simulate drawing samples (civil conflict events) from their underlying distribution. After creating a grid for the country, we will assume that for a given square i, j a civil conflict event:

$$Y_{i,j} \sim \text{Poi}(\mu_{i,j}) \tag{1}$$

We can then place a prior on our rate $\mu_{i,j}$ given our knowledge about other conflicts and other regions (so as to not use the data we are training and testing on to create our prior). In a poisson model, the rate parameter is equal to the mean $\lambda_{i,j} = \mu_{i,j}$. This rate parameter is connected to a linear combination of the predictors $\mathbf{X}$ as follows where β is a set of weighting coefficients:

$$\lambda_{i,j} = \exp(x_{i,j}\beta) \tag{2}$$

This becomes bayesian when we start to consider a prior distribution over these coefficients, β. In practice, the prior is given by the multivariate normal [3]:

$$\beta \sim \mathcal{N}(b_0, B_0^{-1}) \tag{3}$$

where b_0 is the vector of means of the explanatory variables and B_0 is the precision matrix for the explanatory variables.

We can then sample potential future conflicts to get an idea of how these conflicts occur within a region. We will then use the samples from this distribution to generate test cases for the optimization problem that we are concerned with.

3 Optimizing Resource Placement and Allocation

Traveling Salesman Problem

Problem statement

One question of particular interest that we can tackle using techniques from the class is how to route emergency aid to locations where it is needed. It is all well and good to use our *predictive* methods to anticipate where conflicts may occur, but decision makers in a real world scenario would also desire *prescriptive* tools to help carry out their mission. For concreteness, let's postulate a Red Cross medical or food supply caravan that originates from the organization's in-country headquarters. This caravan wishes to visit all n emergent locations in order to deliver needed supplies. They wish to do so in the most efficient manner possible.

Mathematical specification

This is fundamentally an optimization problem, and one that is well known. Here is the traditional convex optimization specification of the problem:

$$\min \sum_{i=0}^{n} \sum_{j \neq i, j=0}^{n} c_{ij} x_{ij}$$

$$\text{s.t.}$$

$$x_{ij} \in \{0, 1\} \qquad\qquad i, j = 0, \cdots, n$$

$$\sum_{i=0, i \neq j}^{n} x_{ij} = 1 \qquad\qquad j = 0, \cdots, n$$

$$\sum_{j=0, j \neq i}^{n} x_{ij} = 1 \qquad\qquad i = 0, \cdots, n$$

$$u_i - u_j + n x_{ij} \leq n - 1 \qquad\qquad 1 \leq i \neq j \leq n$$

This should be immediately recognizable as the traveling salesman problem (TSP), an integer linear program (ILP) where:

- x_{ij} is a binary decision variable indicating whether we go from location i to location j.
- c_{ij} is the distance between location i and location j.
- The objective function is the sum of the distances for routes that we decide to take.
- The constraints ensure that all locations are visited once and only once.

The problem, of course, is that brute force solution of the TSP is $\mathcal{O}(n!)$. Traditional, deterministic algorithm approaches such as branch-and-bound or branch-and-cut are still impractical for larger numbers of nodes. In many cases, exhaustive search for global optimality is not even particularly helpful as long as the solution found is good enough. We will use simulated annealing (SA) to get acceptable solutions to the TSP (c.f. the class lectures and homework problem).

Proposed approach (*The Prestige*)

Hold on to your butts, that's not all. Because we did something similar in class, we will complicate this problem a bit. Consider a mash-up of two NP-hard problems: the traveling salesman problem and the knapsack problem. "What? You guys are crazy." Maybe, but we're doing it live.

We will make the problem double-plus-NP-hard by making this a multi-objective optimization problem where *the contents of the aid trucks* also have an optimization component. Therein lies the knapsack problem: subject to a volume or weight constraint, and given that different locations might have very different needs such as food, vaccinations, or emergent medical supplies, *which supplies do we*
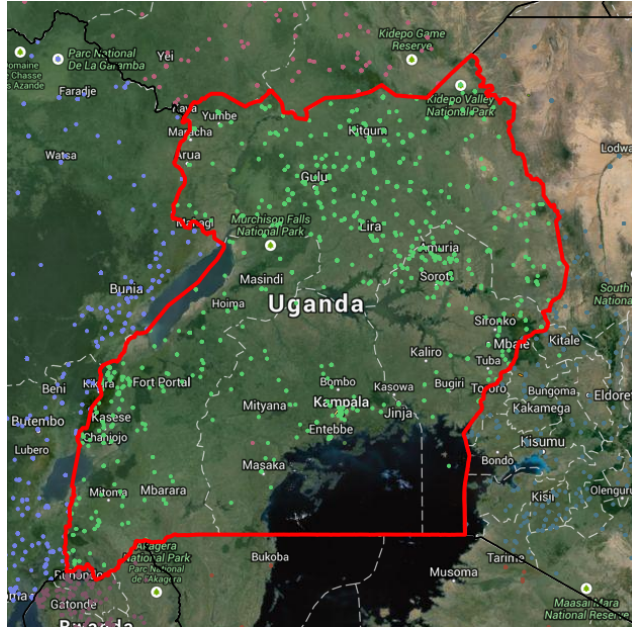
Figure 2: Civil conflicts in Uganda 1997-2013. [2]

pack on the trucks? This may be a toy problem, but if so it is an inherently unsafe toy like a BB gun — you could put your eye out.

Here's the unbounded[1] version of the knapsack problem:

$$\max \sum_{i=1}^{n} v_i x_i$$
$$\text{s.t.}$$
$$x_i \in \mathbb{Z}$$
$$x_i \geq 0$$
$$\sum_{i=1}^{n} w_i x_i \leq W$$

In this formulation:

- x_i is a zero or positive integer decision variable indicating how many units of item i we load on the truck.
- v_i is the utility we get from bringing along item i.
- w_i is the weight of item i.
- W is the maximum weight the truck can carry.

Moreover, if time permits, we will also explore some other stochastic algorithms often used to optimize the TSP including Tabu search (a sort of randomized branch-and-bound) and ant colony optimization (based on the biological analogy of ant-like random walks with attractive properties for walks that are better than others).

---

[1]Often, this problem is formulated such that you can only bring one of each item, but that doesn't make sense here. We want to be able to bring as many types of each type of aid as we think necessary, and we'll assume that as many as desired are available to load on the trucks before starting out from HQ.

Given a set of simulation data about conflicts (along with historical conflicts), we can now use stochastic optimization techniques to determine how to distribute aid resources. Our objective function will attempt to model areas that have: (1) a lower risk of violence, (2) a larger surrounding population (to serve more refugees), and (3) easy access for aid organizations from neighboring countries. We can then use simulated annealing to optimize over the delivery and placement of food, water, medical resources and shelter to the affected populations.

Literature Review

We have reviewed the relevant literature on this problem and found some promising sources. First of all, a similar effort was made by Sebastian Schutte in his 2010 master's thesis. In it, he used the same ACLED data set in order to predict the occurence of violent conflicts in several African nations. While helpful for us as background information, his focus was on the political science aspect. It seems that he used Monte Carlo methods for simulation, but the methods are not described and our sense is that it was probably specified as a problem for BUGS's black box.

Additionally, we found a helpful paper from the Journal of Statistics Education describing this data set and its potential didactic use for class projects. It was helpful in its description of interesting questions that made be answered by using this ACLED data in conjunction with certain other sources.

Finally, we have reviewed the relevant sections of a number of canonical sources for problems such as these including Gelman and Rubin, Gill, and Bishop.

References

[1] Armed Conflict Location and Event Data Project, http://www.acleddata.com/data/version-4-data-1997-2013/.

[2] Map of civil conflicts in Uganda. http://worldmap.harvard.edu/africamap/

[3] Bayesian Poisson Regression (as part of the Zelig package in R) http://people.iq.harvard.edu/ fali-madh/inst/doc/poisson.bayes.pdf