Restless Bandits visiting Villages: A Preliminary Study on distributing Public Health Services

Biswarup Bhattacharya
University of Southern California
Los Angeles, California, USA
bbhattac@usc.edu

ABSTRACT

Distributing public health services is a major challenge. Health workers are responsible for spreading awareness about preventable health problems among the general public by physically visiting the at-risk individuals. However, a limited number of health workers are often responsible for a large population with a variety of health problems. It is therefore essential to design an effective policy to maximize the coverage and spread of health information with limited resources.

In this paper, we propose a novel hierarchical visitation policy design, scalable to regions of various sizes and diversities, which consists of two levels of planning: i) Macro-level planning (region-level) by adapting the p-functional regions problem (PFRP) to our setting, and ii) Micro-level planning (village-level) by formulating a restless multi-armed bandit (RMAB) model and using POMDPs with Whittle Index Policy. We also consider how to address the heterogeneity of health problems across villages to ensure better service delivery and the dynamic nature of public health priorities, which have not been attempted in previous literature, to the best of our knowledge. Our preliminary experiments show promising results which demonstrate the potential of this methodology to be applied for health policy planning.

CCS CONCEPTS

• **Applied computing → Health care information systems**; • **Computing methodologies → Multi-agent planning**; **Partially-observable Markov decision processes**;

KEYWORDS

Public health; p-functional regions problem; Restless multi-armed bandits; Whittle Index policy; POMDP

ACM Reference Format:

Biswarup Bhattacharya. 2018. Restless Bandits visiting Villages: A Preliminary Study on distributing Public Health Services. In *COMPASS '18: ACM SIGCAS Conference on Computing and Sustainable Societies (COMPASS), June 20–22, 2018, Menlo Park and San Jose, CA, USA.* ACM, New York, NY, USA, 8 pages. https://doi.org/10.1145/3209811.3209865

COMPASS '18, June 20–22, 2018, Menlo Park and San Jose, CA, USA

ACM ISBN 978-1-4503-5816-3/18/06…$15.00
https://doi.org/10.1145/3209811.3209865

1 INTRODUCTION

In many developing nations, public health workers are responsible for spreading awareness of health issues in rural areas, usually via patient visits or through health camps. However, there are very limited number of health workers: i.e., there is only an average of 1 health worker per 500 individuals in India [5]. Optimal utilization of these human resources is therefore especially important in raising awareness about health.

In India, over 70% people live in rural areas, where preventable illnesses are a major health issue. Across the country, over 70% of infants are malnourished, infant mortality rate is high, only 15% of mothers receive antenatal care, and mental health issues are prevalent in the population [22]. Studies in Bihar, a region in India, found that there remains a vast lack of health awareness and education in the villages; specifically in the areas of nutrition, immunization, diseases prevention and treatment, maternity care, and family planning [7]. Health workers can address these issues through information dissemination campaigns, which raise awareness about healthy habits, available immunization facilities, health camps, diagnostic camps, and treatment resources [9]. In this paper, we focus on this critical public health problem, and propose an algorithm for conducting optimal information dissemination campaigns both at the regional-level (macro-level) and the village-level (micro-level).

Currently, the Indian health system is organized into regional health centers, each with multiple villages in its jurisdiction. We redistrict these centers in a novel way to improve authorities' abilities to provide oversight, manage logistics, and monitor regional health status, which in turn leads to better resource utilization. This macro-level planning can be repeated every few years as demographics shift over time.

It is also important to look at the dynamics of individual villages to provide regular targeted support. We model this visitation problem of health workers visiting villages as a restless multi-armed bandit problem (RMAB) to distribute k health workers to n villages, and plan the visits using POMDPs. Using the Whittle Index policy, we find the optimal actions for health workers in every round to maximize coverage and information spread. We consider how to match health workers with medical specialties to the specific health needs of each village while determining this visitation policy. This has not been attempted in the previous literature.

We evaluated our model using publicly available Indian Census data. Remarkably, using this algorithm, we estimate that there is potential for health workers to reach over a hundred thousand vulnerable people who would otherwise have never been visited given the current lack of availability of decision aids to assist in village visitation.

2 MACRO-LEVEL PLANNING: DISTRICTING LARGE REGIONS

The first step in our policy design is breaking down large regions into smaller manageable regions, or "public health districts". This districting can work in parallel with the existing administrative divisions in a country. This enables multiple authorities to have a stake in a single public health district, leading to better expected outcomes. All activities in the region can be monitored from a functional center, called the district headquarters.

2.1 Model

Our districting model is heavily inspired by the *p-functional Regions Problem* (PFRP) described in geographical journal literature [4, 13]. We define a functional region to be a region where there is an agglomeration of public health services at a regional center that focuses on providing maximum support to people in the surrounding regions [1]. Formally, the PFRP problem is the identification of a set of p groups that are aggregated into n areal units (public health districts) while optimizing a predefined objective function with a given set of criteria or constraints. The objective function can be formulated to minimize the dissimilarity of areal units or, in our case, to maximize their similarity in public health issues facing each areal unit. In our case, the smallest areal or spatial unit is a village (or a town).

2.2 Optimization Problem

Each of the n villages (basic spatial units) is denoted by $i \in \{1, \ldots, n\}$. Index k denotes a village selected as a functional center of the region. The total number of regions to be districted is p. The set N_i denotes the set of villages adjacent to village i. The following decision variables are used in this model: x_{ik} (whether village i is included in region k), s_{ik} (whether village i is chosen as a sink of region k), and f_{ijk} (the amount of unit flow from village i to village j in region k).

Finally, c_{ik} indicates the magnitude of the actual health-related interactions between the individual villages and the regional centers, such as commuting to clinics and hospitals, to determine the functional regions, whereas f_{ijk} denotes the conceptual flows (health-related movement) between basic villages in a contiguous region. Based on this notation and the corresponding decision variables, the PFRP is modeled as follows:

$$\text{Maximize} \sum_k \sum_i c_{ik} x_{ik} \tag{1}$$

Subject to:

$$\sum_k x_{ik} = 1, \forall i, \tag{2}$$

$$s_{ik} \leq x_{ik}, \forall i, k, \tag{3}$$

$$\sum_i s_{ik} \leq 1, \forall k, \tag{4}$$

$$\sum_k \sum_i s_{ik} = p, \tag{5}$$

$$f_{ijk} \leq x_{ik}(n-p), \forall i, j \in N_i, k, \tag{6}$$

$$f_{ijk} \leq x_{jk}(n-p), \forall i, j \in N_i, k, \tag{7}$$

$$\sum_{j \in N_i} f_{ijk} - \sum_{j \in N_i} f_{jik} \geq x_{ik} - (n-p)s_{ik}, \forall i, k, \tag{8}$$

$$x_{ik} = \{0, 1\}, \forall i, k, \tag{9}$$

$$s_{ik} = \{0, 1\}, \forall i, k, \tag{10}$$

$$f_{ijk} \geq 0, \forall i, j \in N_i, k, \tag{11}$$

$$x_{ij} = 0, \forall i, j \in Q, \tag{12}$$

$$s_{ij} = 0, \forall i, j \in Q, \tag{13}$$

$$x_{ii} = 0, \forall i \in M, \tag{14}$$

$$s_{ii} = 0, \forall i \in M. \tag{15}$$

The objective function (1) maximizes the interactions between the villages and the functional center within a region. The PFRP determines the functional centers and delineates their sphere of influence simultaneously. Constraint (2) ensures that every village i is in only one region k, and constraint (3) ensures that village i can be a sink in a region k only when it is assigned to the same region. Constraints (4) and (5) restrict the number of regions to p by stating that there can be only one sink in each region, with the total number of sinks being p ($p < n$). Constraints (6) and (7) ensure that unit flows can exist only between two adjacent villages in the same region. Constraint (8) establishes the contiguity requirement for regionalization — if a village is not a sink, it must contribute at least one unit flow; and if a village is a sink, then it can have a nonnegative net flow up to $n-p-1$ (largest number of villages that can be assigned to a region). Constraints (9), (10) and (11) impose integer and non-negativity restrictions on the decision variables.

Constraints (12)-(15) are introduced by the ATR method. The PFRP method does not scale well to larger maps, therefore the analytical target reduction (ATR) method is applied. It utilizes the fact that one can search for the optimal p by going q steps at a time, by considering the solution at level $p+q$ (thus, q is defined to be the ATR step size, $1 \leq q \leq p-1 < n$). As described in [13], ATR proceeds by the following steps. Initially, the $p = (n-q)$ case is solved. We can start with all the basic areal units (villages) as the potential seeds or we can start at an intermediate level of p, i.e., with a high q value. Then, unselected villages are added to Q, and the constraints (12) and (13) restrict these villages to be centers and sinks for the next ATR case. Also, it is observed that some villages are repeatedly chosen as functional centers regardless of the p level because of the substantial level of activity (interactions with other units) and we denote these villages as M. A village with dominant inflows from other units can be regarded as a functional center. Thus, applying constraints (14) and (15) allows the solution space can be reduced and the computational efficiency to be improved by considering one or two villages as centers regardless of the p level.

It must be noted that functional centers may not be identical to sinks. The functional center of a region is selected such that it maximizes spatial interactions from other villages i within a region k. It can intuitively mean a village with a good hospital or health center (e.g. a government hospital) where geographical flows from other centers are highly concentrated.

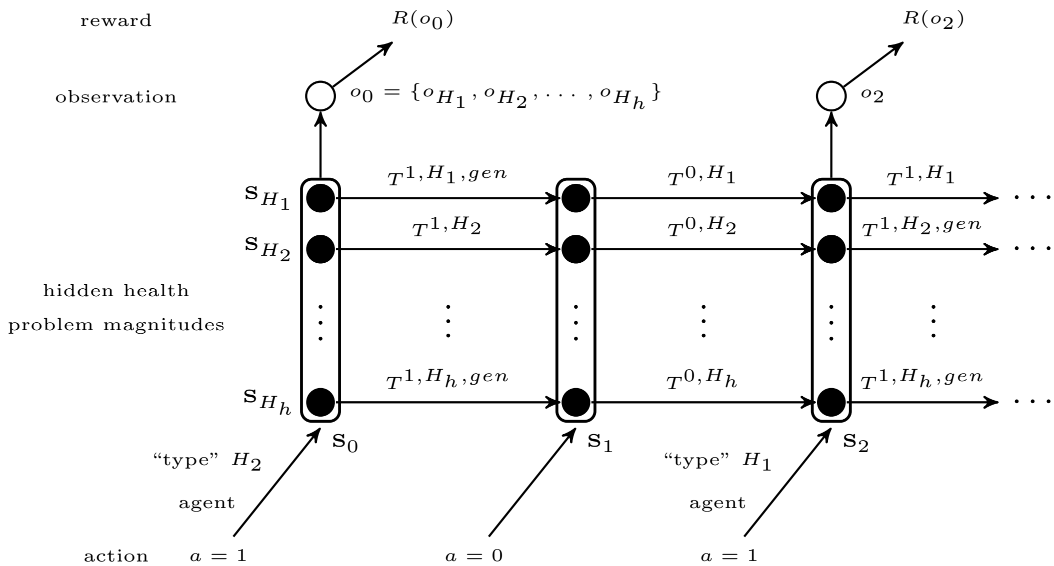


Figure 1: Complete Model

3 MICRO-LEVEL PLANNING: VILLAGE VISITATION

After determining the regions and the grouping of the villages, regular planning of visits by the health workers to the villages is required so that targeted health services can be distributed. Here, the "health worker" term is a proxy for not only actual individual health workers but also health units and health camps set up by teams of public health workers. Henceforth, the term "health agent" will be used to signify a unit of public health resource (individual health workers, camps, mobile units etc.).

We are considering that each "health agent" specializes in certain health procedures and have general non-specialized expertise in disseminating other health information. We also consider that not all villages have the same problems to the same degree. When a village is not visited for a long time, the health situation deteriorates. When a visit is made, even then the health agents do not have perfect information about the health situation and can only make approximate observations given the large number of households in every village. Thus, the question becomes whether to visit new villages for unknown rewards or to keep visiting already visited villages. Simultaneous consideration of health deterioration, medical specialization, and uncertain village needs have not been attempted in previous work, to the best of our knowledge.

We model this problem as a restless multi-armed bandit problem and use the Whittle index policy to identify the optimal strategy. The modeling is heavily inspired by similar strategies applied in wildlife conservation domains, especially in patrol planning of defenders against poachers [6, 19, 26].

3.1 Formal Model & Setup

For every health problem H_i, the problem is to select k out of n villages to visit. The objective is to maximize coverage of affected people and and provide the *required* help at the *correct* locations. The variables are denoted as follows. Let all the health problems be denoted by $H = \{H_1, \ldots, H_h\}$. There are a total of W health agents with k_{H_i} agents for each H_i type of health problem, where $k_{H_i} < n$.

Each village has an associated hidden health problem intensity $\mathbf{S}_{H_i} \in \{0, 1, \ldots, n_s - 1\}$ (higher $\mathbf{S}_{H_i}$ implies higher prevalence of that health problem in the village). However, the health agents observe these intensities according to $\mathbf{O}_{H_i} \in \{0, 1, \ldots, n_o - 1\}$. Note that health agents will make an observation of *all* health problems in a village being visited, but will not be able to make any observation of a village not visited. Thus for every village, the observation is a vector of h elements pertaining to a value from each of $\mathbf{O}_{H_i}$ for all i of that village.

When a village is visited by a health agent of "type" H_d, the hidden health problem magnitude $\mathbf{S}_{H_d}$ transitions according to a $n_s \times n_s$ transition matrix T^{1,H_d}. However, the other hidden health problem magnitudes ($\mathbf{S}_{H_i} \forall i \in \{1, \ldots, h\} \setminus \{d\}$) in the same visited village transitions according to a $n_s \times n_s$ transition matrix $T^{1,H_i,gen}$. If the village is not visited, the hidden health problem magnitude transitions according to a $n_s \times n_s$ transition matrix T^{0,H_i}. The distinction in transition matrices arises from the fact that when a visit is made by a health agent of "type" H_i, they will be able to work on problem H_i more effectively than the other problems. The "gen" matrices are relevant only in case of visited cases; in case of no visit, all health problems deteriorate according to their natural rate. Each health problem has different rates of growth and spread leading to different transition matrices for all health problems. T^1 tends to reduce magnitudes of health problems and T^0 tends to increase these magnitudes.

The health agents not only make observations while visiting a village but also receive rewards for performing activities like immunization, campaigning, etc. We define the reward function $\sum_{i=1}^{h} R(o_i), o_i \in \mathbf{O}_{H_i}$ - larger o_i leads to higher reward $R(o_i)$, as the reward is dependent on the observation made. Note that health agents get h rewards for observing all the h health problems only in the visited village.

In summary, when a health agent of "type" H_d visit a village, she makes an observation regarding every health problem h depending on the current hidden health problem magnitude in that village, gets the reward associated with the observation, and then the hidden health problem magnitude of H_d transitions according to T^{1,H_d} and rest of the hidden problem magnitudes $H_i,\ \forall i \in \{1,\ldots,h\} \setminus \{d\}$ transition according to $T^{1,H_i,gen}$; for the villages the health agents do not visit, they do not have any observation, get reward 0, and the hidden health problem magnitude transitions according to $T^{0,H_i},\ \forall i$ (Figure 1).

While the state discretization level n_s, observation discretization level n_o and reward function $R(o)$ are pre-specified by the public health administration, the transition matrices T^{1,H_i} and T^{0,H_i}, and initial belief can be learned from previous health reports.

4 VISITS BY RESTLESS BANDITS

In the multi-armed bandit problem, there are n arms out of which k arms need to be activated at every round. Each arm represents an independent Markov machine. The states of the active arms transition after every round, while the passive arms remain in the same state as before. However, in RMABs, the passive arms also transition at every round. It is PSPACE-hard to find the optimal strategy to general RMABs [17], therefore index policies are used which assigns a value to each arm to measure how rewarding it is to activate an arm at every stage. In literature, Whittle index is used for RMABs [24]. Whittle Index is based on the concept of providing enough subsidy to every arm which would make passive action optimal for the current state. Whittle index policy chooses to activate the k arms with the highest Whittle indices. Larger m implies larger gap between active action (activate) and passive action, and therefore it is more attractive the player to activate this arm. Formally, let $V_m(x; a = 0)$ be the maximum cumulative reward the player can achieve until the end if he takes passive action at the first round at the state x with subsidy m, and $V_m(x; a = 1)$ be the maximum cumulative reward the player can achieve until the end if he takes active action at the first round at the state x with subsidy m. Whittle index $I(x)$ of state x is then defined to be:

$$I(x) \triangleq \inf_m \{m : V_m(x; a = 0) \geq V_m(x; a = 1)\}$$

An arm is indexable if $\phi(m)$ monotonically increases from $\emptyset$ to the whole state space as m increases from $-\infty$ to $+\infty$. An RMAB is indexable if every arm is indexable. Here, $\phi(m)$ is the set of states for which passive action is optimal given subsidy m.

4.1 Restless Bandit Formulation

The health agents do not have perfect knowledge of the health state of every village, hence they maintain a belief b_{H_i} of each possible state of every possible health problem (H_i) in every village, based on which decisions are taken. The belief update occurs by Bayesian rules. The updated belief state for every health problem H_i is hereby described when the health agents visit a village ($a = 1$), or do not visit a village ($a = 0$). When the health agent who visits the village is of "type" H_d, the belief update performed is as follows (where $d \in \{1,\ldots,h\}$):

$$b'_{H_i}(s') = \begin{cases} \eta_1 \sum_{s \in S_{H_i}} b_{H_i}(s)\, O^{H_i}_{so}\, T^{1,H_i}_{ss'}, & a = 1, i = d \\ \eta_2 \sum_{s \in S_{H_i}} b_{H_i}(s)\, O^{H_i}_{so}\, T^{1,H_i,gen}_{ss'}, & a = 1, i \neq d \\ \sum_{s \in S_{H_i}} b_{H_i}(s)\, T^{0,H_i}_{ss'}, & a = 0. \end{cases} \tag{16}$$

where η_1, η_2 are the normalization factors. The belief update is intuitive — when a visit is made ($a = 1$), the belief first updates according to the observation made and then the transition happens according to T^{1,H_i} or $T^{1,H_i,gen}$ using the logic as described previously; and when a visit is not made, there is no observation, and the transition happens according to T^{0,H_i}. The following mathematical relations, pertaining to the Whittle index policy, can now be written for our problem (here $i \in \{1,\ldots,h\}$):

$$V_m(b_{H_i}; a = 0) = m + \beta V_m(b_{H_i, a=0}) \tag{17}$$

$$\begin{aligned} V_m(b_{H_i}; a = 1) = &\sum_{s \in S_{H_i}} b_{H_i}(s) \sum_{i=1}^{h} \sum_{o \in O_{H_i}} O^{H_i}_{so}\, R(o) \\ &+ \beta \sum_{o \in O_{H_i}} \sum_{s \in S_{H_i}} b_{H_i}(s)\, O^{H_i}_{so}\, V_m(b^o_{H_i, a=1}) \end{aligned} \tag{18}$$

$$V_m(b_{H_i}) = \max\{V_m(b_{H_i}; a = 0), V_m(b_{H_i}; a = 1)\} \tag{19}$$

$$I(b_{H_i}) \triangleq \inf_m \{m : V_m(b_{H_i}; a = 0) \geq V_m(b_{H_i}; a = 1)\} \tag{20}$$

$$\phi_{H_i}(m) \triangleq \{b_{H_i} : V_m(b_{H_i}; a = 0) \geq V_m(b_{H_i}; a = 1)\} \tag{21}$$

Equation 17 specifies that when a village is not visited, the immediate reward received is the subsidy and there is a β-discounted future reward. The value function $V_m(b_{H_i,a=0})$ is updated from b_{H_i} using the case $a = 0$ in Equation 16. Equation 18 specifies that when a village is visited, health agents get the expected immediate reward (first term) and there is a β-discounted future reward. $V_m(b^o_{H_i,a=1})$ is the value function at new belief $b^o_{H_i,a=1}$ that is updated from b_{H_i} according to Equation 16, case $a = 1$ with observation o. Equation 19 is the final value function. Equation 20 specifies the Whittle Index for belief b_{H_i} and Equation 21 specifies the passive action set $\phi_{H_i}(m)$, which is the set of belief states for which passive action ("not visit") is the optimal action given subsidy m.

4.2 Numerical Evaluation of Whittle Index

One can prove sufficient conditions for indexability, however due to the already existing rich literature on indexability proofs for RMABs [8, 14–16], it is skipped in this paper. For other problems, we numerically evaluate their indexability.

Proposition 1. *If $m < hR(0) - \beta h \frac{R(n_o-1)-R(0)}{1-\beta}$, $\phi(m) = \emptyset$; if $m > hR(n_o - 1)$, $\phi(m)$ is the whole belief space.*

Proof. If $m < hR(0) - \beta h \frac{R(n_o-1)-R(0)}{1-\beta}$, let $V_m(b; a = 0) = m + \beta P$ and $V_m(b; a = 1) = \sum_{i=1}^{h} R(o_i) + \beta Q$. In the worst case, $P \leq \frac{hR(n_o-1)}{1-\beta}$ by getting $hR(n_o - 1)$ reward in every round; similarly $Q \geq \frac{hR(0)}{1-\beta}$ by getting $hR(0)$ reward in every round. $\sum_{i=1}^{h} R(o_i) \geq hR(0)$, hence we get $V_m(b; a = 1) - V_m(b; a = 0) = \sum_{i=1}^{h} R(o_i) - m + \beta(Q - P) \geq hR(0) - m + \beta h \frac{R(n_o-1)-R(0)}{1-\beta} > 0$. Thus, being active is always the optimal action for any state so that $\phi(m) = \emptyset$.

If $m > hR(n_o - 1)$, then the strategy of always being passive dominates other strategies so $\phi(m)$ is the whole belief state space. □

It can also be proven that $\phi(m)$ increases monotonically for $m \subseteq [hR(0) - \beta h\frac{R(n_o-1)-R(0)}{1-\beta}, hR(n_o - 1)]$. Therefore the RMAB is indexable for m in this particular range as specified by the definition of indexability. Given the subsidy m, $\phi(m)$ can be determined by solving a POMDP model. Given the indexability, the Whittle Index can be found by simply doing a search within the range $m \subseteq [hR(0) - \beta h\frac{R(n_o-1)-R(0)}{1-\beta}, hR(n_o - 1)]$, described in detail in [19].

4.3 Planning with POMDPs

The passive action set $\phi(m)$ is computed using a POMDP model whose conditional observation probability is dependent on the start state and action, unlike in standard POMDPs where the conditional observation probability is dependent on end state and action [12] (Figure 2).

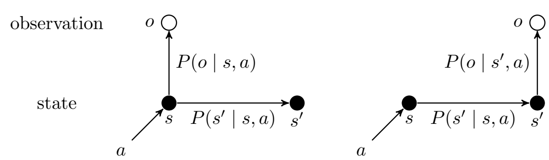


Figure 2: Special POMDP (left) v/s Standard POMDP (right)

Every single health problem H_y in every village is modeled as a POMDP model. The POMDP is formulated as follows:

- **State space**: The state space is a vector $\mathbf{S} = \{\mathbf{S}_{H_1}, \ldots, \mathbf{S}_{H_h}\}$, where the space for each of $\mathbf{S}_{H_y}$ is $\{0, 1, \ldots, n_s - 1\}$.
- **Action space**: The action space is $A = \{0, 1\}$, where $a = 0$ represents passive action (do not visit) and $a = 1$ represents active action (visit).
- **Observation space**: The observation space is a vector $\mathbf{O} = \{\mathbf{O}_{H_1}, \ldots, \mathbf{O}_{H_h}\}$, where the space for each of $\mathbf{O}_{H_y}$ is $\{-1, 0, 1, \ldots, n_o - 1\}$. The $o = -1, o \in \mathbf{O}_{H_y}$ is used to represent no observation when taking action $a = 0$. When health agents take action $a = 1$, they may make observations $\mathbf{O}_{H_y} \setminus \{-1\}$.
- **Conditional transition probability**: The conditional transition probability is as follows, when $a = 1$ implies a health agent of "type" H_d visits the village.

$$P(s'_{H_y} \mid s_{H_y}, a) =$$

$$\begin{cases} P(s'_{H_y} = j \mid s_{H_y} = i, a = 1, y = d) = T_{ij}^{1,H_y} \\ P(s'_{H_y} = j \mid s_{H_y} = i, a = 1, y \neq d) = T_{ij}^{1,H_y,gen} \\ P(s'_{H_y} = j \mid s_{H_y} = i, a = 0, y \in \{1, \ldots, h\}) = T_{ij}^{0,H_y}. \end{cases}$$

- **Conditional observation probability**:

$$P(o_{H_y} \mid s_{H_y}, a) =$$

$$\begin{cases} P(o_{H_y} = j \mid s_{H_y} = i, a = 1) = O_{ij}^{H_y}, \\ P(o_{H_y} = -1 \mid s_{H_y}, a = 0) = 1, \qquad \forall s \in \mathbf{S}_{H_y}. \end{cases}$$

- **Reward function**:

$$R(s_{H_y}, s'_{H_y}, a, o_{H_y}) = \begin{cases} 0, & a = 0, \\ \sum_{y=1}^{h} R(o_{H_y}), & a = 1. \end{cases}$$

The computation of $R(s_{H_y}, a)$ is done as follows:

$$R(s_{H_y}, a) = \sum_{s' \in \mathbf{S}_{H_i}} P(s'_{H_y} \mid s_{H_y}, a) \times \sum_{i=1}^{h} \sum_{o \in \mathbf{O}_{H_i}} P(o_{H_y} \mid s_{H_y}, a)\, R(s_{H_y}, s'_{H_y}, a, o_{H_y})$$

For value iteration, the belief update equation is slightly different from standard POMDPs [19].

4.4 Complete POMDP

We can combine the POMDP models of every village to form a full POMDP model which describes our problem. Such h complete POMDPs can be created to consider all the health problems.

- **State space**: $\mathbf{S} = \mathbf{S}^1 \times \mathbf{S}^2 \times \cdots \times \mathbf{S}^n$. Denote $s = (s^1, s^2, \ldots, s^n)$.
- **Action space**: $A = \{(a^1, a^2, \ldots, a^n) \mid a^j \in \{0, 1\}, \forall j \in \mathbb{N}, \sum_{j \in \mathbb{N}} a^j = k\}$ (as there are only k health agents of type H_i. Denote $a = (a^1, a^2, \ldots, a^n)$.
- **Observation space**: $\mathbf{O} = \mathbf{O}^1 \times \mathbf{O}^2 \times \cdots \times \mathbf{O}^n$. Denote $o = (o^1, o^2, \ldots, o^n)$.
- **Conditional transition probability**: The probability is $P(s' \mid s, a) = \prod_{j \in \mathbb{N}} P^j(s'^j \mid s^j, a^j)$.
- **Conditional observation probability**: The probability is $P(o \mid s, a) = \prod_{j \in \mathbb{N}} P^j(o^j \mid s^j, a^j)$.
- **Reward function**: $R(s, s', a, o) = \sum_{j \in \mathbb{N}} R(s^j, s'^j, a^j, o^j)$.

5 EMERGENCY INTERVENTION STRATEGY

A "warning score" S_i for every village which signifies the overall health situation of a village can be maintained. It is calculated based on the current beliefs of every state $(b_{i,H_j}(s))$ of every health problem H_j in village i.

$$S_i = \sum_{j=1}^{h} \sum_{s \in \mathbf{S}^i_{H_j}} s b_{i,H_j}(s)$$

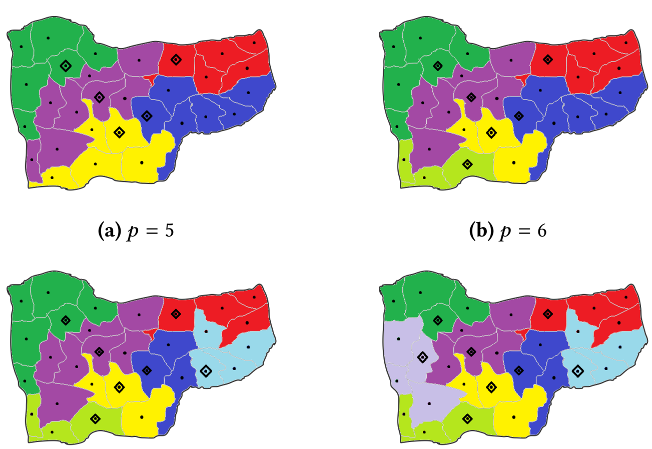
If the S_i goes above a threshold, it is an indication to the authorities that more comprehensive health measures need to taken in that village, and it may be required to send health agents of multiple specialties. The maximum possible value of $S_i = h(n_s - 1)$, and that occurs when every health problem has reached the maximum intensity (catastrophic levels) in that village. This score can be utilized to notify village-level or region-level health emergencies or epidemics.

6 EXPERIMENTAL EVALUATION

We consider a state in India — Bihar, one of the most disadvantaged states in terms of the health situation and health delivery systems. Bihar consists of 38 districts and around 45000 revenue villages with a total rural population of 91 million people. We consider a region of 30 villages (population 60,000) in one of the districts in Bihar — Arwal (population 700,000) for our experimental evaluation. The types of health agents working in this district include Accredited Social Health Activists (ASHAs), Auxiliary Nurse Midwives (ANMs), among others. Due to lack of exact data about the number of health workers in Arwal, we estimate from statistics [20] that this district is served by roughly 1 health worker per 1700 individuals. There are 5 Primary Health Centres and 65 Health Subcentres in Arwal. These coordinate the activities of the health agents. The major health problems in this district include lack of infant care and ante-natal care, malaria, TB, and leprosy.

We use demographic data from the Indian Census 2011 [10]. The data available includes information on population demographics, literacy, working class information, etc., at the village level. Villages close to each other generally face similar health problems due to similarity in demographics and development. So, we can segregate the large region into sub-regions which will enable better planning as we will have better estimation of the problems in sub-regions than in the whole region. Once these new "public health districts" are created, health agents can be deployed sub-region by sub-region.

We combined the population demographics and working class information as inputs to the optimization problem presented in macro-planning. Taking each village as the smallest spatial unit, we run the algorithm and identify the functional centers and the number of "public health districts" formed. Given the available data, we took the population in every village as the indicator for activity. As public health is a resource utilized by all people of all ages, population is a reasonable proxy given no other suitable alternatives.



(a) $p = 5$ (b) $p = 6$ (c) $p = 7$ (d) $p = 8$

Figure 3: Illustration of optimal districting of 30 villages for different number of regions (p-regions)

Figure 3 shows the optimal solutions of a region consisting of 30 villages for $p = 5$ to 8. The lines in the figure demarcate the villages purely for ease of illustration. The black dots are the village centers. The colors display the regions formed, and the diamonds are the functional centers selected after optimization. It is observed that the administrative seats are matching with functional centers. It is interesting to note that having a fully-functioning government hospital can indicate higher chance of the village being set as a sink. It is clear that with even more appropriate features and data, this districting can be done with higher accuracy. The authorities do have access to such data but we do not, hence we can expect that in the real world this algorithm will perform better than our simulation.

Thus, for framing the visitation policy, we say that these colored regions are the "public health districts" in this region and a central health center needs to be active at each of the functional centers. Health workers can be posted to these functional centers, where micro-level planning will be conducted to plan the visitation policies for each of these health workers to visit every village as and when necessary. The optimization can be run hierarchically to further divide districts and so on, till the desired level of regionalization is achieved. We notice in our model simulated with real data that the number of villages in each region is approximately 5 to 6 which implies that it is ideal if micro-level planning is conducted in batches of 5 to 6 villages.

The micro-level planning performance is evaluated in terms of the cumulative reward achieved within the first 30 rounds ($\beta = 0.9$). All results are averaged over 1000 simulation runs.

6.1 Evaluation of Whittle Index Policy

We will compare the Whittle Index policy (WI) with the following baselines:

(1) **Random**: Randomly allocate k health agents of a certain "type" in every round.
(2) **Myopic**: Allocate k health agents of a certain "type" to the villages with the highest immediate reward in every round.
(3) **Exact POMDP** (POMDP): Plan using the exact POMDP model discussed. The exact algorithm used for value iteration is the incremental pruning algorithm [2].
(4) **POMCP**: Plan using the POMCP algorithm to solve the exact POMDP model discussed. POMCP algorithm uses a particle filter to maintain an approximation of the belief state and uses MCTS to simulate the next step to find the best action [21]. It is especially useful when POMDP becomes infeasible to solve for large state spaces.

Table 1: Solution Quality for small-scale problem: $n = 2, k = 1, h = 2, n_s = n_o = 2$

Random	Myopic	POMCP	POMDP	WI
3.280	3.521	3.590	3.717	3.695

The solution quality values in Table 1 indicate rewards obtained, roughly translating to, in the real world, number of individuals positively affected by the health workers. As expected, exact POMDP provides the highest solution quality. However, Whittle Index policy performs reasonably well — comparable to exact POMDP in this case and much better than POMCP or Myopic. It must be noted that although exact POMDP provides the best solution quality, it is not scalable.

Table 2: Evaluation for $n = 5, h = 2, n_s = n_o = 2$, varying k

k	Random	Myopic	POMCP	WI
1	7.441	10.512	11.810	12.124
2	10.431	14.030	15.281	16.425
3	14.129	18.459	18.780	19.294

From Table 2, it is clear that with the increase in number of health agents, the cumulative reward increases, with Whittle Index policy performing very well in finding rewarding strategies.

Table 3: Evaluation for $n = 5, h = 3, n_s = n_o = 2$, varying k

k	Random	Myopic	POMCP	WI
1	10.441	16.002	18.111	19.356
2	15.006	20.141	22.164	23.379
3	21.463	23.988	25.665	26.414

Table 3 shows the same problem with an added health problem. The cumulative rewards increase as a whole due to the additional observations being made about the added health problem. It was noted that after implementing macro-planning, the number of villages in the problem reduced (to subproblems) thereby giving a boost in performance.

It is important to understand the consequences of the results obtained. The reward function is a function of the observations which are related to the intensities of the health problems. The highest number in the observed intensity scale ($n_o - 1$) indicates that the majority of the people in the village suffers from the problem. Using the Census data, we can understand the significance of the results obtained. The average population of a village in India is around 2000 [10]. The results in Table 2 show a cumulative reward of 26.414 for WI which roughly translates to the expected number of people reached to around 1750, whereas a reward of 23.988 for Myopic translates to around 1600 people reached. Scaling this to a large region of 30 villages, for example, as in the macro-level planning experiment, corresponds to WI policy reaching approximately 1000 more people, who would otherwise have never been reached. On a national scale, this number can potentially translate to at least hundreds of thousands of people reached due to the Whittle Index policy, as there are over 500,000 villages in India.

6.2 Evaluation of RMAB Model

The RMAB model was compared with other existing algorithms — UCB, EXP3, and SWUCB. The performance is evaluated in terms of the cumulative reward achieved within the first 30 rounds ($\beta = 0.9$) after multiple learning rounds.

Table 4: RMAB evaluation & comparison for varying learning rounds (LR) for $n_s = n_o = 2$

LR	Random	EXP3	SWUCB	UCB	RMAB
500	4.411	5.334	5.500	6.271	6.534
1000	4.421	5.536	5.742	6.534	6.849
1500	4.417	5.711	5.818	6.669	7.110

The superior performance of RMAB becomes even more pronounced as the number of learning rounds increases as illustrated in Table 4.

7 RELATED WORK

Population health planning and region-based planning has been explored previously in healthcare literature. [11] notes that regionalization is supposed to have facilitated a better alignment between the allocation of healthcare resources and population health needs. In our approach, informed regionalization is taking place during macro-planning, which should improve the results when applied in the real world.

Restless multi-armed bandits (RMAB) is a well-studied model in literature. However, to the best of our knowledge, the RMAB model has previously never been attempted to be used in the public health domain. However, bandit theory has been explored previously in medical literature. In [25] and [18], they propose bandit-based designs for attempting clinical trials. Optimal design of clinical trials was attempted using RMABs and Whittle Index Policy in [23].

Planning visitation in villages can be thought of to be similar to multi-agent patrolling policies which have been researched previously. In [19], they plan patrols using RMAB in forests to tackle poachers. However, our work deals with public health and the complexity of our problem is much larger due to the variety of health problems which can exist even at one location. The concept is similar, but the detailed settings are different. We also propose a hierarchical model, which is especially relevant in the domain of public health. In [3], they consider the problem of patrolling under uncertainty and threats by solving a single-agent POMDP model.

8 CONCLUSION

We have presented a hierarchical model — PFRP method for macro-planning and a RMAB approach with Whittle Index Policy for micro-level planning. Even though the results obtained are only after preliminary analysis, we believe similar approaches can be adopted by health administrations for planning health policies in the future. It is important to test and fine-tune this algorithm on real world data, and it is possible that there can be multiple types of complications due to the diversity of the real world.

9 ACKNOWLEDGMENT

Biswarup Bhattacharya is partially supported by the Annenberg Graduate Fellowship at the University of Southern California, Los Angeles, CA, USA.

REFERENCES

[1] Lawrence A Brown and John Holmes. 1971. The delimitation of functional regions, nodal regions, and hierarchies by functional distance approaches. *Journal of Regional Science* 11, 1 (1971), 57–72.

[2] Anthony Cassandra, Michael L Littman, and Nevin L Zhang. 1997. Incremental pruning: A simple, fast, exact method for partially observable Markov decision processes. In *Proceedings of the Thirteenth conference on Uncertainty in artificial intelligence*. Morgan Kaufmann Publishers Inc., 54–61.

[3] Shaofei Chen, Feng Wu, Lincheng Shen, Jing Chen, and Sarvapali D Ramchurn. 2015. Multi-agent patrolling under uncertainty and threats. *PloS one* 10, 6 (2015), e0130154.

[4] Juan C Duque, Luc Anselin, and Sergio J Rey. 2012. The max-p-regions problem. *Journal of Regional Science* 52, 3 (2012), 397–419.

[5] Victoria Fan and Sudhir Anand. 2016. The Health Workforce in India. (2016).

[6] Fei Fang, Peter Stone, and Milind Tambe. 2015. When Security Games Go Green: Designing Defender Strategies to Prevent Poaching and Illegal Fishing. In *IJCAI*. 2589–2595.

[7] MS Friedman and Jyoti Somani. 2002. Health conditions in the tribal villages of South Bihar: an epidemiologic survey. *The Journal of the Association of Physicians of India* 50 (2002), 1376–1380.

[8] KD Glazebrook, D Ruiz-Hernandez, and C Kirkbride. 2006. Some indexable families of restless bandit problems. *Advances in Applied Probability* 38, 3 (2006), 643–672.

[9] Mahaveer Golechha. 2016. Health promotion methods for smoking prevention and cessation: A comprehensive review of effectiveness and the way forward. *International journal of preventive medicine* 7 (2016).

[10] Government of India. 2011. Data.gov.in. `http://data.gov.in/`. Accessed: 2017-10-24.

[11] Jeremiah Hurley. 2004. Regionalization and the allocation of healthcare resources to meet population health needs. *Healthcare Papers* 5 (2004), 34–39.

[12] Leslie Pack Kaelbling, Michael L Littman, and Anthony R Cassandra. 1998. Planning and acting in partially observable stochastic domains. *Artificial intelligence* 101, 1 (1998), 99–134.

[13] Hyun Kim, Yongwan Chun, and Kamyoung Kim. 2015. Delimitation of Functional Regions Using ap-Regions Problem Approach. *International Regional Science Review* 38, 3 (2015), 235–263.

[14] Jerome Le Ny, Munther Dahleh, and Eric Feron. 2008. Multi-UAV dynamic routing with partial observations using restless bandit allocation indices. In *American Control Conference, 2008*. IEEE, 4220–4225.

[15] Keqin Liu and Qing Zhao. 2010. Indexability of restless bandit problems and optimality of whittle index for dynamic multichannel access. *IEEE Transactions on Information Theory* 56, 11 (2010), 5547–5567.

[16] Jose Nino-Mora. 2001. Restless bandits, partial conservation laws and indexability. *Advances in Applied Probability* 33, 1 (2001), 76–98.

[17] Christos H Papadimitriou and John N Tsitsiklis. 1999. The complexity of optimal queuing network control. *Mathematics of Operations Research* 24, 2 (1999), 293–305.

[18] William H Press. 2009. Bandit solutions provide unified ethical models for randomized clinical trials and comparative effectiveness research. *Proceedings of the National Academy of Sciences* 106, 52 (2009), 22387–22392.

[19] Yundi Qian, Chao Zhang, Bhaskar Krishnamachari, and Milind Tambe. 2016. Restless poachers: Handling exploration-exploitation tradeoffs in security domains. In *Proceedings of the 2016 International Conference on Autonomous Agents & Multiagent Systems*. International Foundation for Autonomous Agents and Multiagent Systems, 123–131.

[20] Krishna D Rao, Aarushi Bhatnagar, and Peter Berman. 2012. So many, yet few: human resources for health in India. *Human resources for health* 10, 1 (2012), 19.

[21] David Silver and Joel Veness. 2010. Monte-Carlo planning in large POMDPs. In *Advances in neural information processing systems*. 2164–2172.

[22] B Venkatashiva Reddy, Arti Gupta, Ayush Lohiya, and Pradip Kharya. 2013. Mental health issues and challenges in India: A review. *International Journal of Science and Research Publications* 3, 2 (2013), 1–8.

[23] Sofía S Villar, Jack Bowden, and James Wason. 2015. Multi-armed bandit models for the optimal design of clinical trials: benefits and challenges. *Statistical science: a review journal of the Institute of Mathematical Statistics* 30, 2 (2015), 199.

[24] Peter Whittle. 1988. Restless bandits: Activity allocation in a changing world. *Journal of applied probability* 25, A (1988), 287–298.

[25] S Faye Williamson, Peter Jacko, Sofía S Villar, and Thomas Jaki. 2017. A Bayesian adaptive design for clinical trials in rare diseases. *Computational statistics & data analysis* 113 (2017), 136–153.

[26] Rong Yang, Benjamin Ford, Milind Tambe, and Andrew Lemieux. 2014. Adaptive resource allocation for wildlife protection against illegal poachers. In *Proceedings of the 2014 international conference on Autonomous agents and multi-agent systems*. International Foundation for Autonomous Agents and Multiagent Systems, 453–460.