International Journal of Research in Engineering, Science and Management
Volume-3, Issue-7, July-2020
journals.resaim.com/ijresm | ISSN (Online): 2581-5792

Food Distribution During Flood Using Q-Learning Algorithm

Mohammad Ashfan[1], Albin Francis[2], T. R. Deekshith[3*], Ameen Ahmed[4], Sayeesh[5]
[1,2,3,4]*Student, Department of Computer Science and Engineering, Alva's Institute of Engineering and Technology, Moodbidre, India*
[5]*Professor, Department of Computer Science and Engineering, Alva's Institute of Engineering and Technology, Moodbidre, India*
*Corresponding author: deekshithtr11@gmail.com

***Abstract*: During Emergencies, when we want to distribute food to different localities it is not always possible to use airways. In such cases when we have no other option other than roadways, we can implement Q-Learning to reach the localities. In this algorithm, the localities act as nodes or vertices and the routes acts as edges. A satellite image is fed to the system of the flood affected area. Starting from the first node, it decides the action to be performed in each stage as it proceeds. Solving, a valid solution would need to represent a route where every location is included at least once and only once. If a route contains a single location more than once, or missed out a location completely then it wouldn't be valid.**

***Keywords*: Disaster, Emergency, Food, Flood, Q-learning, Rescue, Shortest path.**

1. Introduction

Natural disasters have always posed a critical threat to human beings, often being accompanied by major loss of life and property damage. In recent years, we have witnessed more frequent and intense natural disasters all over the world. At the times of emergencies, such as floods when we want to distribute food to different localities it is not always possible to use airways. In such cases when we have no other option other than roadways, we can implement Q Learning to reach the localities. In this algorithm, the localities act as nodes or vertices and the routes acts as edges. A satellite image is fed to the system of the flood affected area. Starting from the first node, it decides the action to be performed in each stage as it proceeds. Solving, a valid solution would need to represent a route where every location is included at least once and only once. The adverse impacts of a disaster can be substantially miti-gated if during the disaster accurate information regarding the available volunteers can be gathered and victims' locations can be determined in a timely manner, enabling a well-coordinated and efficient response. This is particularly apparent whenever there is a huge burst of requests for limited public resources. Since the phone line resource is limited, many phone calls did not get through and victims turned to social media to plead for help, posting requests with their addresses.

At the same time, many willing volunteers seeking to offer help during the disaster were left idle as no one knew where they should be sent. In the case of a hurricane, a major challenge is that without coordination, multiple volunteers with boats may go to rescue the same victim while other victims have to wait for extended times to be rescued. This mismatch between victims and volunteers represents an enormous waste of limited volunteer resources.

It is therefore imperative to improve the emergency services coordination to enable them to efficiently share information, coordinate rescue efforts and allocate resources more effectively, and offer guidance for optimal resource allocation.

The problem of resource coordination has drawn considerable attention in the computer science community, and several data mining frameworks have been developed to address this problem. Previous researchers have primarily focused on three approaches: supervised learning, adaptive methods, and optimization-based method. Traditional supervised learning models demand a dataset that is statistically large in order to train a reliable model for example by building regression models to predict needs and schedule resources accordingly.

2. Related Work

Different methods have been developed to solve the vehicle routing problem (VRP) and vehicle routing problem with time windows (VRPTW). The routes should be chosen to minimize the total distribution cost. A detailed description of these and other related problems including a literature review of the methods are provided. The solution is presented by one or more routes, each associated to one vehicle. Each customer must be assigned only to one vehicle and the load must not exceed the vehicle capacity. Various works address the transportation of foodstuffs, dealing with pertinent issues to our problem.

1. Tarantilis and Kiranoudis (2011) analyzed the distribution of the fresh milk. They formulated the problem as a heterogeneous fixed fleet vehicle routing problem; this is a VRP with vehicles that have different capacity. A threshold-accepting based algorithm was developed aiming to satisfy the distribution needs of the company, allowing them to schedule their distribution many times a week.

2. Tarantilis and Kiranoudis (2012) presented a real-life distribution problem of fresh meat in an area of the city of Athens. They formulated the problem as an open multi depot vehicle routing problem. They presented a new stochastic search meta-heuristic algorithm belonging to the class of threshold-accepting algorithms.
3. It Hwang (2015) presented an effective distribution model for determining optimal patterns of food supply and inventory allocation for famine relief areas. He modeled a VRP that incorporated inventory allocation and optimal distribution based on minimizing the deprivation and starving instead of travel distance or time.
4. Prindezis, Kiranoudis, and Marinos- Kouris (2016) presented an application service provider, to be used for central food markets, which coordinates and disseminates tasks and related information for solving the VRP. For the solution of the VRP they used a metaheuristic technique based on the tabu search. They tailored their software to the road network of Athens and applied it to the integrated-logistics problem of deliveries to the 690 retail companies that comprise the Athens Central Food Market. They used a two-phase algorithm to the VRP. In the first phase, a route construction algorithm was used and in the second phase a tabu search was used to improve the given solution.
5. Faulin (2017) presented the implementation of the mixed algorithm procedure that uses heuristic and exact subroutines in the solution of a VRP having specific constraints related to companies in the agribusiness field.
6. Gendreau, and Potvin (2018) proposed a time-dependent model for the VRPTW. The model that they developed is based on time-dependent travel speeds and satisfies the first-in-first-out (FIFO) property. They extended the tabu search heuristic to solve the problem and showed that the time-dependent model provides substantial improvements over a model based on fixed travel-times.

3. Problem Formulation

In recent times, due to different factors the earth platelets have become irregular and also because of global warming, the melting of ice hills and glaciers led to floods in different places. It has become difficult to predict in advance the area which could be affected by these calamities in advance. Also the unpredictable rain has become a key factor for floods and inadequate measures the prevent it and act when such calamities occurs. In such areas affected by flood, shortage or unreachability of resource becomes a major problem. Our system is dedicated to supply food to such areas food in large scale and shortest time. The problem of resource coordination has drawn considerable attention in the computer science community, and several data mining frameworks have been developed to address this problem. Previous researchers have primarily focused on three approaches: supervised learning, adaptive methods, and optimization-based method. Traditional supervised learning models demand a dataset that is statistically large in order to train a reliable model, for example by building regression models to predict needs and schedule resources. Unfortunately, due to the unique nature of resource management for disaster relief, it is generally impractical to model this using traditional supervised learning models. Every disaster is unique and hence it makes no sense to model one disaster relief problem by using the dataset collected from other disasters;

A realistic dataset for that disaster can only be obtained when it occurs. This means that traditional supervised learning is unable to solve the highly individual resource management problems associated with disaster relief efforts. Other researchers have developed adaptive methods and proposed adaptive systems for resource allocation. However, a common limitation of the adaptive approach is that the parameters in adaptive models change slowly and hence converge slowly.

An alternative is to model resource coordination problems as simulation problems or optimization problems which requires the process of modelling and tuning repeatedly if any of the external environmental parameters change.

Real world resource coordinating problems are very challenging for a number of reasons:

1. The sample size is small, especially in the early stages of the disaster, when there is almost no available data. Any decision-support system needs to move fast and make decisions swiftly.
2. The real-world environment where the resource coordination actually happens is a highly complex system with multiple uncertainties. For instance, the locations of volunteers and victims are dynamically changing, and the rescue time for an arbitrary victim varies depending on factors such as traffic, road closures, and emergency medical care, many of which are also changing dynamically.
3. There is no well-defined objective function to model the scheduling problem for disasters, especially when victims need emergency care or collaborative rescue efforts.

The recent success achieved in applying machine learning to challenging decision-making domains suggests that Reinforcement Learning (RL) is a promising method with, reinforcement learning has been successfully applied to solve problems such as optimizing deep neural networks with asynchronous gradient descents for the controllers One appealing feature of the reinforcement learning method is that it can overcome many of the difficulties involved in building accurate models, which is usually formidable given the scale and complexity of real- world problems. Moreover, reinforcement learning does not require any prior knowledge of system behavior to learn optimal strategies. This means that reinforcement learning can be used to model systems that include changes and/or uncertainties. Finally, reinforcement learning can be trained for objectives that are hard to optimize directly because of the lack of precise models. When reward signals that are correlated with the objective are involved, this can be modelled by reinforcement learning as It is possible to incorporate a variety of goals by adopting different reinforcement rewards. We build an efficient heuristic multi-agent reinforcement learning framework for large-scale disaster rescue work based on information gathered by mining social media data. This study is one of the first that specifically focuses on coordinating volunteers in disaster relief using reinforcement learning. As information comes in about volunteers and victims' situations and makes recommendations to minimize the total distance travelled by all the volunteers to supply the maximum possible number of victims.

4. Proposed Method

Q-Learning Algorithm is implied in which each node is visited at least once starting from the starting node which is initialized as Q. An action is chosen which in this case is measuring the distance from the nodes which are not visited. The action is performed and a reward is measure and granted.

Algorithm 1: ResQ in Rescue Scheduling

let t=0, Q_t^i=1;
initialize s_0;
repeat
  Observe current state S_t;
  A_t = HeuristicActionSelection(S_t)
  Every volunteer execute its action a_t^i in A_t;
  Observe $R_t^i \ldots R_t$ and $a_t^i \ldots t$

$$Q_{t+1}^i(s, a^1 \ldots a^N) = (1 - a_t)Q^i(s, a^1 \ldots a^N) + a_t\{r_t^i + \gamma\pi^i(s_{t+1}) \sum_{j=1}^{N} Q_t^j(s_{t+1})\pi^j(s_{t+1})\}$$

  where $(\pi^i(s_{t+1}), \pi^j(s_{t+1}))$ are cooperative strategies;
  Let t=t+1;
until *rescue complete*;

Once it is done for that particular node, Q value is updated and procedure is repeated until all the nodes are evaluated. The satellite image is fed to the system implementing Q-Learner algorithm. The image is read in by the Symbolic Learner to draw the nodes.

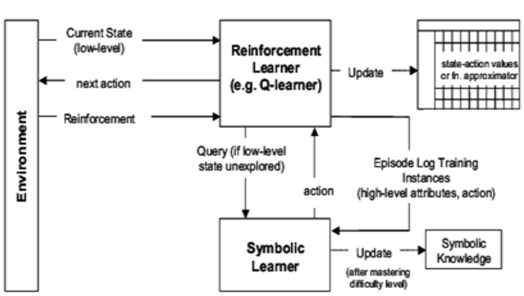


Fig. 1. Block diagram

Once the nodes are plotted the data is passed back to the Reinforcement learner. After each stage actions are updated and approximation is done and reward is calculated and assigned.

The ResQ algorithm is applied as shown above. Heuristic Action Selection function is applied in ResQ to calculate the minimum distance and update the action Heuristic action selector.

Algorithm 2: Heuristic action selection

function HeuristicActionSelection (S_t)
Input : State S_t
Output: best $found\_action$
Choose best actions A based on policy $\pi(S_t)$ and Q
$min\_distance = \infty$
for $action_n \in A$ **do**
  $next\_state_n = perform\_actions(action\_n)$
  $distance$ = HeuristicDistance($next\_state_n$)
  if $distance \leq min\_distance$ **then**
    $min\_distance = distance$
    $found\_action = action_n$
  end
end
return $found\_action$

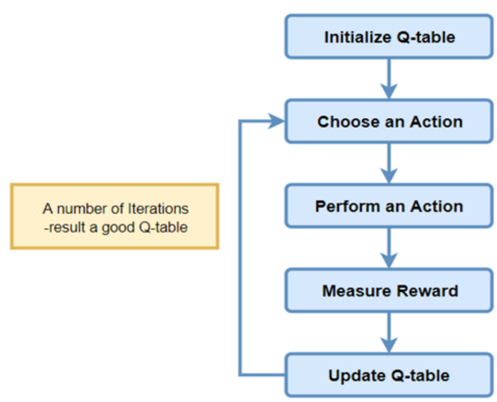
The minimum distance is calculated for every state and is compared with distance of previous state. The minimum distance is recorded for each stage and corresponding action is taken.

The drawbacks in the existing system were mainly due to the repeating visit of some localities and in case if the routes were supposedly long, they were often skipped. We propose a system using Q-Learning Algorithm in which each node is visited at least once starting from the starting node which is initialized as Q. An action is chosen which in this case is measuring the distance from the nodes which are not visited. The action is performed and a reward is measure and granted. Once it is done for that particular node, Q value is updated and procedure is repeated until all the nodes are evaluated.

Step 1: Initialize the Q-Table

First the Q-table has to be built. There are n columns, where n= number of actions. There are m rows, where m= number of states. In our example n=Go Left, Go Right, Go Up and Go Down and m= Start, Idle, Correct Path, Wrong Path and End. First, let's initialize the values at 0.

Step 2: Choose an Action



Step 3: Perform an Action

The combination of steps 2 and 3 is performed for an undefined amount of time. These steps runs until the time training is stopped, or when the training loop stopped as defined in the code.

First, an action (a) in the state (s) is chosen based on the Q-Table. Note that, as mentioned earlier, when the episode initially starts, every Q-value should be 0.

Then, update the Q-values for being at the start and moving right using the Bellman equation.

Steps 4: Measure Reward

Now we have taken an action and observed an outcome and reward.

Steps 5: Evaluate

We need to update the function Q(s,a).

This process is repeated again and again until the learning is stopped. In this way the Q-Table is been updated and the value function Q is maximized. Here the Q(state, action) returns the expected future reward of that action at that state.

5. Conclusion

we have proposed a food distribution system using Q-Learning algorithm which helps for predicting the best possible routes. This helps the rescue team to reach out the victim's location by using the shortest possible route driven by the algorithm.

References

[1] Long H. Nguyen, “Coordinating Disaster Emergency Response with Heuristic Reinforcement Learning.”
[2] Saravi S, Kalawsky R, Joannou D, Rivas-Casado M, Fu G and Meng F., Use of artificial intelligence to improve resilience and preparedness against adverse flood events, Water (Switzerland), Volume 11, Issue 5, 2019, Article No. 973.
[3] Heung-Suk Hwang, “A food distribution model for famine relief.”
[4] Ana Osvald, Lidija Zadnik Stirn, “A vehicle routing algorithm for the distribution of fresh vegetables and similar perishable food,” Journal of Food Engineering, Volume 85, Issue 2, 2008, pp. 285-295.