A New Optimization Approach to Resource Distribution using Semi-Supervised Learning Graphs

Mubarak Banisakher — Mubarak.banisakher@saintleo.edu
Derek Mohammed — derek.mohammed@saintleo.edu
Van Nguyen — Van.nguyen@saintleo.edu

Department of Computer Science
Saint Leo University, FL. USA

Abstract **- An effective Post-Disaster Management System (PDMS) will result in distribution of emergency resources such as, hospital, storage and transportation in a reasonable and equitable manner. This study starts with semi-supervised learning (SSL) based graph system to provide post-disaster path optimizations. Next, the graph-based resource is converted to a directed graph resulting in an adjacent matrix. Decision information is provided in two ways: clustering operation and graph-based semi-supervised optimization. The PDMS in this study incorporates a path optimization algorithm based on Ant Colony Optimization (ACO) that results in cost-effective resource distribution. Simulation results demonstrate the effectiveness of the proposed methodology by comparing ACO with clustering based algorithms of tour improvement algorithm (TIA) and Min-Max Ant System (MMAS).**

Keywords - Data, Graph, Algorithms, Post-disaster, Optimization, networks, resources

I. INTRODUCTION

This paper presents a post-disaster management system (PDMS) that will enable effective and equitable distribution of emergency resources such as hospitals, storage and transportation in the affected areas. We apply semi-supervised learning (SSL) based graphs to PDMS to convert to a directed graph resulting in an adjacent matrix. Based on the matrix, a decision is enabled by the PDMS in two ways: the first one is clustering operations, and the second one is the graph-based semi-supervised optimization process. In this study path-optimization algorithm based on ant colony optimization (ACO) was used for minimizing the cost in a post-disaster situation. The methodology is validated through a simulation and the results show that the proposed methodology will be more effective for calculating the optimization path in PDMS.

II. LITERATURE SURVEY AND WEAKNESSES OF CURRENT TECHNIQUES

In the last decade, the global community has experienced an increasing trend in the occurrence and impact of natural and manmade disaster numbers. According to the Center for Research on the Epidemiology of Disasters (CRED), the number of disasters resulting in 100,000 to 999,999 victims around the globe doubled during 1987-2006 (CRED 2006). In 2010, 385 natural disasters were reported worldwide killing more than 297,000 persons, affecting over 217.0 million others and causing US$123.9 billion in economic damages (CRED 2011). Homeland Security News Wire (2012) reports "2011: costliest ever year for earthquakes, weather-related disasters." The 2011 Japan and New Zealand disasters account for two-thirds of estimated $380 billion, with global economic losses nearly two-thirds higher than in 2005, the previous record year with losses of $ 220 billion.

The alarming and devastating impacts of these disasters on human lives and the global economy has resulted in an increased interest in research that focuses on information technology and technical resources in emergency management (National Academies of Science, 2009) . Specifically, there is a need to study the impact that resources such as information and sensor technology can have on the ability to improve logistics, planning and overall responsiveness in emergency management. To be effective, this research must have a human centric approach as the effectiveness of the technology, logistics and information management activities will be driven by the human element of this complex system (Bennett, 2010).

The sensor technology focus of the post-disaster network identifies existing and developing sensing mechanisms, as well as emerging challenges, that are present in the integration of sensor technology throughout the disaster management continuum. A breadth of sensor technology is currently being used to support disasters and new approaches are surfacing on almost a daily basis. These technologies include sensor warning systems, photonics sensors, and broad applications of wire sensing devices and even robotics. (Chebrolu, K. et. al. 2008) As the world's technical capability increases, the development of wireless and satellite sensor technology is likely to grow at a compatible rate. Data driven solutions from advanced technologies such as sensors can aid both in early warning detection and relief logistics during and after a disaster.

A. Semi-Supervised Method

The earliest use of the notion of semi-supervised learning algorithm was mentioned as self-training (Chapelle et al., 2006); this notion also appeared in some of the literature of the 1960s and 1970s (e.g., 1965 Scudder literature) (Scudder, 1995), in Fralick (1967), and in Agrawala (1970).

As the basic idea of the training methods is to use supervised learning techniques to learn the marked data, unlabeled data and then use the result of learning to get it marked, then the new tag data are added to the marked data for the purpose of the learning process. However, the performance of this method depends on the supervised learning technology, and on when to use the 0-1 cost function empirical risk minimization learning, otherwise, unlabeled samples will lapse (Scudder, 1965).

Following a brief introduction of semi-supervised learning, there are two common assumptions that need to be considered:
(1) The clustering hypothesis refers to a sample point in the same cluster, which is likely to have the same category tags. This assumption can be expressed by another equivalent, that is, through the decision boundary regions; which it will be a relatively sparse area of data points. If the decision boundary through the data points is in a relatively dense area, then it is likely to be sample points in a cluster where it is divided into different categories, and clustering contradicts the assumption.

(2) Flow assumption is the existence of low-dimensional characteristics of high-dimensional data and representation of another species is an example of a small local neighborhood having similar properties (Zhou, 2007). The purpose of supervised learning method is to identify things; the identified data with labels are the result of the recognition performance. Training sample set must be composed of a labeled sample, rather than SSL methods that only analyze the data set itself. If data sets show some kind of aggregation, they can be classified according to the nature of aggregation, but it is not with a pre-classification label for this purpose (Pang et al., 2009; Weng et al., 2008; Zhang et al., 2010).

B. Semi-Supervised Learning Process with Graph

Graph-based SSL uses a graph representation of the data with a node for each labeled and unlabeled information. The algorithm is to create an optimization objective function or minimization cost function that consists of a loss function and a regularizer. We have adopted an alternative minimization method for PDMS map calculation introduced by Wang et al. (2008). It starts from a simple function as Gaussian fields and harmonic functions (Zhu et al., 2003).

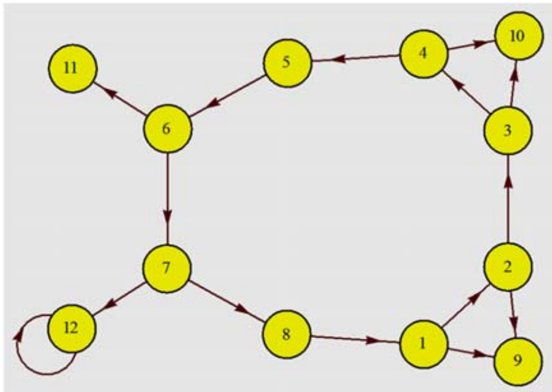
C. Graph-Based Semi-Supervised Learning Algorithm

For each data point to its k nearest neighbors or to examples within some distance ε, the weight W_{ij} of an edge between x_i and x_j is $e^{\frac{-\|x_i-x_j\|^2}{\varepsilon}}$. The minimization problem becomes:

$$\arg\min_{f\in H}\left(\frac{1}{l}\sum_{i=1}^{l}V(f(x_i),y_i)+\lambda_A\|f\|_H^2 \quad \lambda_I\int_M f(x)\|\nabla_M f(x)\|^2\,dp(x)\right) \tag{1}$$

Where ***H*** is a reproducing kernel Hilbert space and ***M*** is the manifold on which the data lie. λ_A and λ_I are the smoothness parameters. The graph is used to approximate the intrinsic regularization term. Defining the graph Laplacian $L = D - W$, we have:

$$f^T Lf = \sum_{i=1}^{l+u}\sum_{j=1}^{l+u} W_{ij}(f_i - f_j)^2 \approx \int_M f(x)\|\nabla_M f(x)\|^2\,dp(x) \tag{2}$$



$$W \times \begin{bmatrix} 0 & 1 & 0 & 0 & 0 & 0 & 0 & -1 & 1 & 0 & 0 & 0 \\ -1 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ . & . & . & . & . & . & . & . & . & . & . & . \\ . & . & . & . & . & . & . & . & . & . & . & . \\ 0 & 0 & 0 & 0 & 0 & 0 & -1 & 0 & 0 & 0 & 0 & 1 \end{bmatrix}$$

Figure 1 Resources Distribution Converted as Graph and Its Adjacent Matrix

There are a number of procedures of scheduling and distribution of emergency supplies in the event of disasters and immediate response process during emergencies. The goal is to ensure an efficient supplies deployment and minimization of loss. Our method focuses on GIS technology to achieve an optimal distribution route for emergency supplies. Path analysis is the most fundamental and most important in GIS functionality, and the core technology is to solve the shortest path by using GIS data in the shortest path algorithms. It is a must to establish the relationship and connections between roads and junctions and abstract them into nodes and arcs in the graph; this process is an analogy to "building the network topology". Topology construction of the road network is needed to achieve the essential foundation for the shortest path algorithm because an efficient implementation of the shortest path algorithm, is critical for the shortest path analysis.

III. A NEW OPTIMIZATION APPROACH

A. Theorem and Definitions

Supposed that $\Omega = (H, R)$, H is the hospital set started from H_0 to end H_{k-1} that means there are k hospitals that need to be assigned emergency resources. $R = \{(i, j) \mid i, j \in H, i \neq j\}$ is the edge set between hospitals (node) H_i and H_j and D_{ij} is the distance between H_i and H_j. We need a shortest path from H_0 to end H_{k-1}.

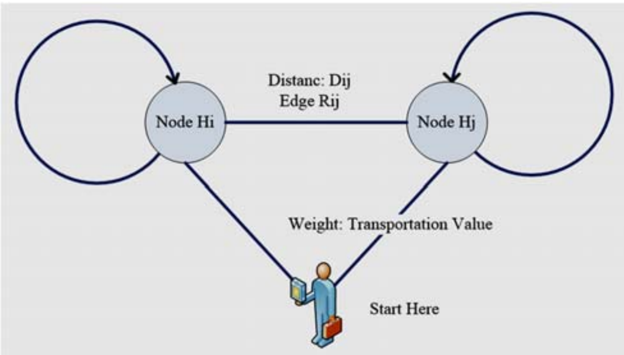


Figure 2 Individuals in Post-disaster System for Emergency Resources Connection

B. Design of the Algorithms

B1 Edge Selection: An ant is a simple computational agent in the ant colony optimization algorithm. It iteratively constructs a solution for the problem at hand. The intermediate solutions are referred to as solution states. At each iteration of the algorithm, each ant moves from a node i to node j, corresponding to a more complete intermediate solution. Thus, each ant k computes a set $B_k(x)$ of feasible expansions to its current state in iteration; and moves to one of these in probability. For ant k, the probability p_{ij}^k of moving from node i to node j depends on the combination of two values, viz., the *attractiveness* η_{ij} of the move, as computed by some heuristic indicating the a priori desirability of that move and the trail level τ_{ij} of the move, indicating how proficient it has been in the past to make that particular move.

The trail level represents a posteriori indication of the desirability of that move. Trails are updated usually when all ants have completed their solution, increasing or decreasing the level of trails corresponding to moves that were part of "good" or "bad" solutions, respectively.

In general, the k-th ant moves from state x to state y with probability:

$$p_{ij}^k = \frac{\tau_{ij}^\alpha \eta_{ij}^\beta}{\sum_{j \in allowed_j} \tau_{ij}^\alpha \eta_{ij}^\beta} \tag{3}$$

where τ_{ij} is the amount of pheromone deposited for transition from node i to j; $\alpha \geq 0$ is a parameter to control the influence of τ_{ij}; η_{ij} is the desirability of state transition ij (*a priori* knowledge, typically $\frac{1}{d_{ij}}$, where d is the distance); and $\beta \geq 1$ is a parameter to control the influence of η_{ij}. τ_{ij} and η_{ij} represent the attractiveness and trail level for the other possible state transitions.

B2. Pheromone Update: When all the ants have completed a solution, the trails are updated by:

$$\tau_{ij} \leftarrow (1-\rho)\tau_{ij} + \sum_k \Delta\tau_{ij}^k \tag{4}$$

where τ_{ij} is the amount of pheromone deposited for a node transition ij; ρ is the pheromone evaporation coefficient; and $\Delta\tau_{ij}^k$ is the amount of pheromone deposited by k-th ant, by:

$$\Delta\tau_{ij}^k = \begin{cases} 1/L^k & the\, k-th \quad ant's\ path \quad arc \\ 0 & others \end{cases} \tag{5}$$

where L_k is the cost of the k- th ant's tour (typically length) and Q is a constant.

B3. Path Optimization for the SSL Graph of Emergency Resource Distribution: 2 sets of rules are employed:

Transmission Rule: In our research here, a method to determine the roulette ant to transfer node between 0 and 1 transition probability random number comparison, when

DOI 10.5013/IJSSST.a.19.04.17

ISSN: 1473-804x online, 1473-8031 print

the transition probability is greater than the random number, it will move to the next point ant *k*.

Update Rules: Ants are in the structure solution process; each moves one step, that of the corresponding arc segment updating pheromone, using local updating rules as follows:

$$\tau_{ij}^{n+1} = \rho_l \tau_{ij}^{n} + (1-\rho_l)\tau_0 \tag{6}$$

where ρ_l represents a single-step move to pheromone persistence, (1- ρ_l) is pheromone volatility. For each arc global pheromone, the update rule is as follows:

$$\tau_{ij}^{n+1} = \rho \tau_{ij}^{n} + \sum_{k=1}^{m} \Delta\tau_{ij}^{k} + \sigma\Delta\tau_{ij}^{*} \tag{7}$$

where

$$\Delta\tau_{ij}^{k} = \begin{cases} 1/L^k & the\, k-th \quad ant's\, path \quad arc \\ 0 & others \end{cases} \tag{8}$$

$$\Delta\tau_{ij}^{*} = \begin{cases} 1/L^* & if\,(i,j) \quad is\, the \quad best \\ 0 & others \end{cases} \tag{9}$$

TABLE 1 PSEUDOCODE OF ACO FOR EMERGENCY RESOURCES DISTRIBUTION BASED ON THE GRAPH

Begin
Initial;
$t \leftarrow 0$;
iteration← 0;
set random m *ants on* n *nodes;*
loop:
set initial start up of each ant in solved set;
for $i \leftarrow 0$ *to* $n-1$ *do*
for $k \leftarrow 1$ *to* m *do*
select node j *follow the probability;*
move ant k *to node* j *;*
set node j *into solution set of ant* k *;*
end for;
$t \leftarrow t+1$;
end for;
count L *objective value of each ant;*
update solution set
count satisfaction of each solution;
$t \leftarrow t+1$;
clear;
iteration← iteration+1;
if iteration< threshold;
then goto Loop;
output solution;
End

We added only one used single ant for updating the pheromone trails after each iteration that was introduced by Stützle and Hoos (2000).

TABLE 2 IMPROVED ACO BY MAX-MIN ANT SYSTEM (MMAS)

Set $\tau_{(i,j)} = \frac{1}{2}$ *for* $all(i,j) \in Graph$
*Construct a solution x**
*Update pheromones w.r.t. x**
Loop:
Construct a solution x
If O(x)>=O(x) then x*:=x*
*Update pheromones w.r.t. x**
Else: Break
Goto Loop

IV. METHODOLOGY AND RESULTS

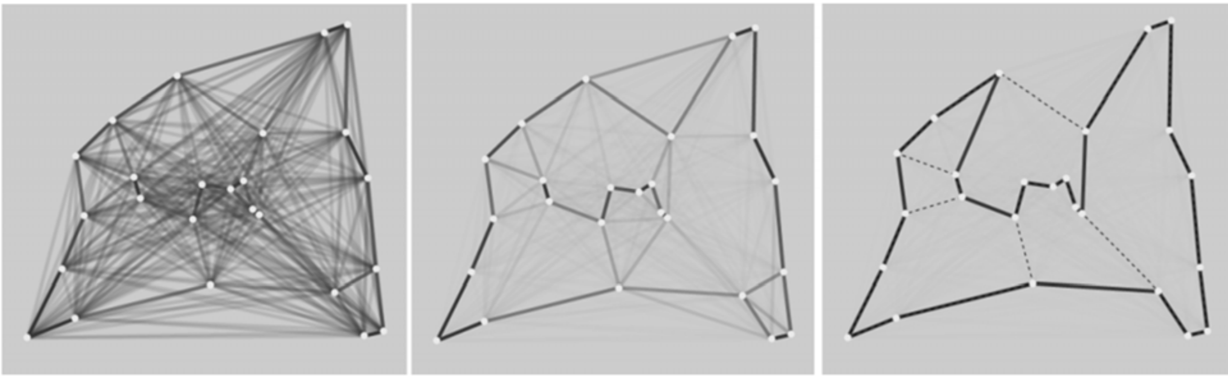
The global best solution deposits pheromone on every iteration along with all the other ants using MMAS. Max-Min pheromone amounts are $[\tau_{\min}, \tau_{\max}]$. All edges are initialized to $\tau_{\max}$ and re-initialized to $\tau_{\max}$ when nearing stagnation (Dorigo, 2007).

Set Hospitals = 50 in PDMS, and Ants = 50, lc is the level of convergence that will be assigned an initial value. The parameters of basic ACO algorithm are listed in Table 3.

TABLE 3 THE PARAMETERS OF BASIC ACO ALGORITHM

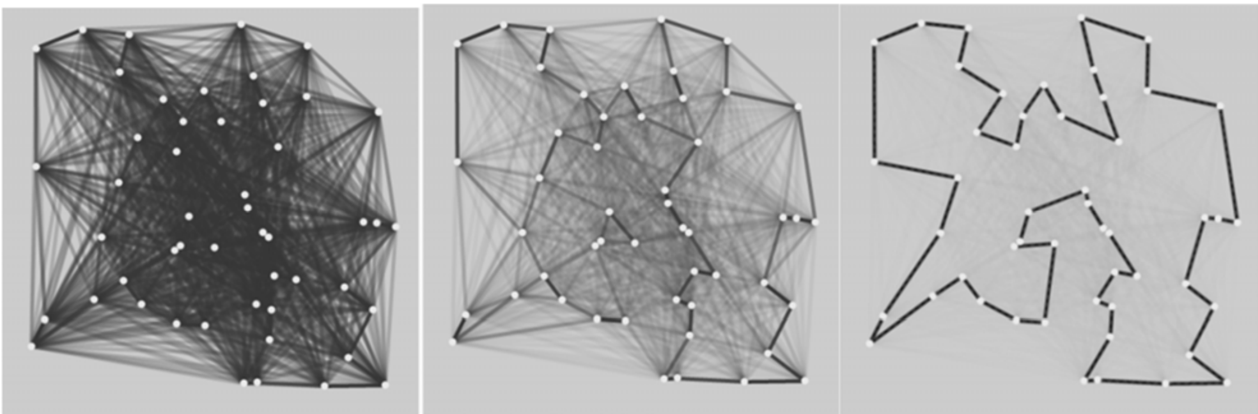
Parameters	Value
Elite	10%
Candidate list	20m
Pheromones: Eva	0.98
Initial level	1.00
Min/Max ratio	0.01
$\tau_{\min}$	1
$\tau_{\max}$	3

The results that only applied TIA are shown in Figure 3. The final tour cost is 26.99. TIA here is based on the nearest neighbor algorithm that was" if there have a 5 node graph, then number the node and start from 1, consider the router 1-2-3-4; and swap the middle nodes to change the order 1-3-2-4; then compare the two and keep the lower weight order. Finally, move to node 2 and repeat until each node has been the start node once."



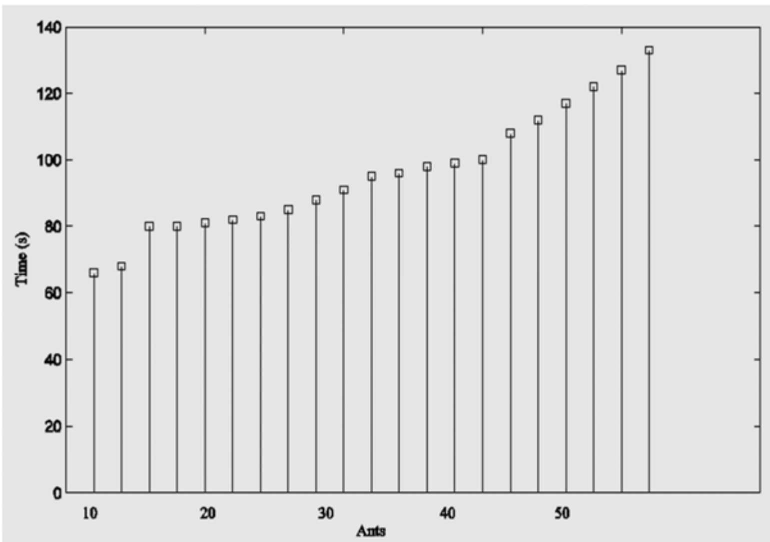
(a) Time=7.7 s, lc=0.07 (b) Time=28.1 s, lc=0.8 (c) Time=147.6 s, lc=0.93
Figure 3 The lc and time values using ACO by TIA only

Set Hospitals = 50 in PDMS, and Ants = 50, lc is the level of convergence that will be assigned an initial value. The results with only TIA and MMAS applied are shown in Figure 4. The final tour cost is 39.23.



(a) Time=87.2 s, lc=0.08 (b) Time=170.4 s, lc=0.1 (c) Time=471.8 s, lc=.93
Figure 4 The lc and time values using ACO by TIA and MMAS

We also compared different hospitals and ants under the same time convergence level. Figure 5 shows the comparison results.



(Ant=10, Time=66 s; Ant=20, Time=82 s; Ant=50 Time=133 s)
Figure 5 Results of Comparison on Population of System under Ants and Time Consumed

By different algorithms and their combinations, we compared the TIA, MMAS and TIA+MMAS, in which N denotes a no-limit start; the dataset is listed in Table 5.

TABLE 5 THE LEVEL OF CONVERGENCE BY DIFFERENT ALGORITHMS

Ants	N	TIA	MMAS	TIA+MMAS*
10	0.44	0.44	0.23	**0.67**
20	0.52	0.56	0.55	**0.7**
30	0.56	0.67	0.6	**0.77**
40	0.6	0.68	0.55	**0.82**
50	0.68	0.78	0.89	**0.98**
60	0.66	0.44	0.67	**0.77**
70	0.56	0.67	0.87	**0.76**
80	0.42	0.53	0.55	**0.56**

(* is the proposed method in this paper)

Figure 6 shows that there is a relation as LC (N)<LC(TIA)<LC(MMAS)<LC(TIA+MMAS) (LC= level of convergence) under node =50, ants<80

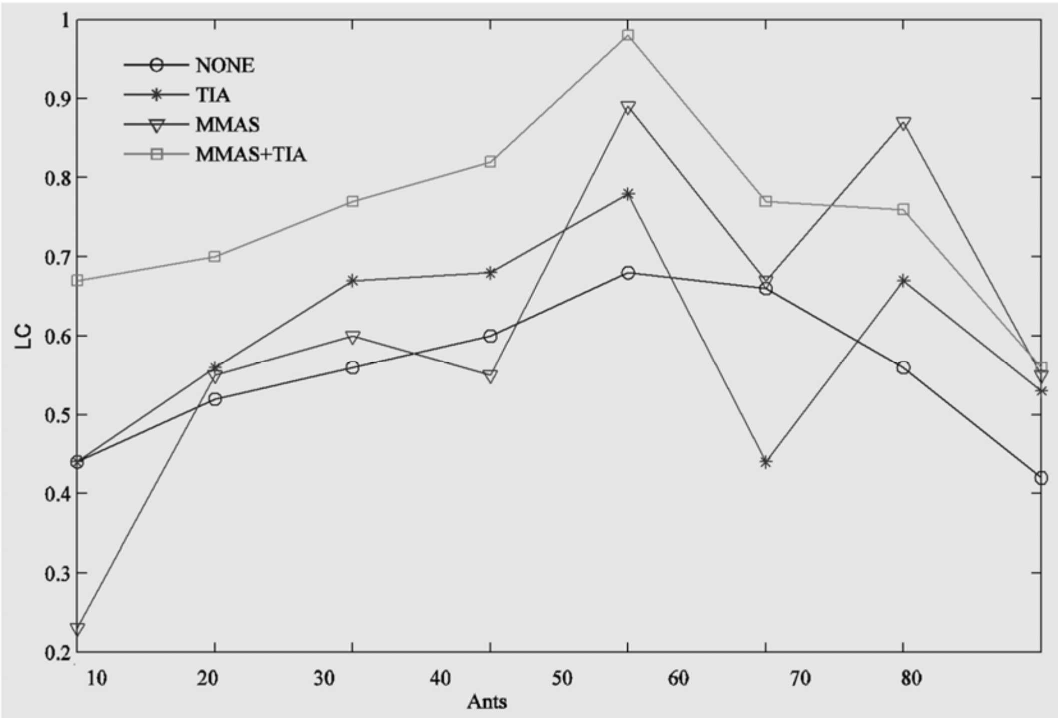


Figure 6 TIA and MMAS of LC Under the Same Population Systems

V. CONCLUSION

Emergency resource distribution is a critical factor in post-disaster management situation. In fact, the distribution effectiveness relies on the map and the nodes' organization, such as the connections and the weights of each node. This paper proposed an effective organized disaster map based on SSL that is more advantageous for optimal path. TIA+MMAS based on the SSL graph shows its effectiveness compared with TIA or MMAS of ACO and other organized maps. Simulation results were also conducted to support our findings and it was as follows:

-1. TIA+MMAS is more effective than TIA or MMAS under different ants and population (nodes).

-2. For the best ants under same nodes, SSL is more effective than FG, KNN, and SOM organized maps.

REFERENCES

[1] Bennett. R (2010), Research and Development, Educational Testing Service, Princeton, NJ, Volume 17, 2010 – Issue 2. Published online: 30 Apr 2010

[2] Chapelle, O., Schölkopf, B., & Zien, A. (2006). Semi-Supervised Learning. Cambridge, MA: The MIT Press, London, England.

[3] CRED (2006) . CRED Labs - Center for Research on Environmental Decisions

[4] CRED (2011) . CRED Labs - Center for Research on Environmental Decisions

[5] Dorigo. M (2007), Ant Coloney Optimization Scholarpedia: URL // www.scholarpedia.org/article/Ant_colony_optimization.

[6] Chebrolu, K. et. al.(2008) a sensor network system for railway bridge monitoring, MobiSys '08 Proceedings of the 6th international conference on Mobile systems, applications, and services Pages 2-14

[7] Pang, Y., Chen J., & Wang, F. (2009). A Special Supervised Learning Algorithm and Its Applications. Hybrid Intelligent Systems, 2009. HIS '09. Paper presented at Ninth International Conference (pp. 462 – 466). IEEE Press.

[8] Scudder, H. J. (1965). Probability of error of some adaptive pattern-recognition machines. IEEE Transactions on Information Theory, 11:363-371.

[9] Scudder, H. J. (1995). Probability of error of some adaptive pattern-recognition machines. IEEE Transactions on Information Theory, 11, 363-371.

[10] Stützle and Hoos (2000). MAX–MIN ant system. Future generation computer systems , Volume 16 Issue 8 Pages 889-914

[11] Wang. J, Jebara. T, and S.-F. Chang, (2008), Graph transduction via alternating minimization, In Proceedings of the Twenty- Fifth International Conference on Machine Learning, Helsinki, Finland, ACM, 2008, 307, 1144-1151.

[12] Weng, P. H., Liu, F. T., Chen, Y. J., et al. (2008). A Hybrid Supervised Neural Network Learning Algorithm. Innovative Computing Information and Control, 2008. Paper presented at ICICIC '08. 3rd International Conference (pp.294-297). IEEE Press.

[13] X. Zhu, Z. Ghahramani, and J. Lafferty (2003), Semi-supervised learning using Gaussian fields and harmonic functions. In Proceedings of the Twentieth International Conference on Machine Learning, Washington DC, USA, AAAI Press, 2003, 912-919.

[14] Zhou Yan-xi, Liu Gang-jun, Fu Er-jiang, et al. (2009), An object-relational prototype of GIS-based disaster database, The 6th International Conference on Mining Science & Technology, Procedia Earth and Planetary Science, 1, 1060-1066.

[15] Zhang, C., Zhao, T., & Li, W. (2010). Automatic search of geospatial features for disaster and emergency management. *International Journal of Applied Earth Observation and G*418.