Machine Learning Approach for Estimating Sensor Deployment Regions on Satellite Images (ISITES 2014)

Enes Ateş, Assoc.Prof.Dr.Aybars Uğur (Ege University)
Asst.Prof.Dr.Tahir Emre Kalaycı (Celal Bayar University)

18 June 2014
Karabuk University, Karabuk, Turkey

Outline

Introduction

- In this study, a fast technique is proposed to estimate the suitable regions for sensor network deployment
- Also sensor count needed to deploy the whole usable area is calculated
- Complementary interactive image processing software is developed
- This work is performed under the TÜBİTAK Project No: 113E947

Wireless Sensor Network

- Consist of small sensor nodes with limited processing and computing resources
- These sensor nodes can:
  - sense
  - measure
  - gather information from the environment
  - transmit the sensed data to the user based on some local decision process
- They are inexpensive compared to traditional sensors

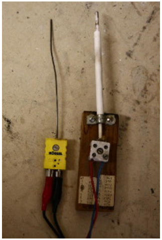

Fig. 1 : Thermocouple sensor for high temperature measurement [6]

Wireless Sensor Network

- WSNs facilitate monitoring and controlling of physical environments from remote locations
- They have great potential for many applications, such as:
  - environmental monitoring
  - military target tracking and surveillance
  - natural disaster relief
  - biomedical health monitoring etc.

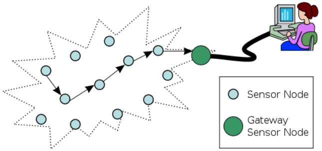

Fig. 2 : Wireless Sensor Network architecture [7]

Sensor Deployment

- Determining the location of the sensors before deploying them to the monitoring area
- There are some optimization problems:
  - monitoring maximum area
  - using minimum number of sensors
  - monitoring different parts of the area that have different priorities

Sensor Deployment Regions

- Different regions (forests, seas, residential areas etc.) of an area can be monitored with WSNs
- If forests will be monitored, mountains, sea or residential areas are irrelevant and no sensors are dedicated to these regions
- Specific deployment can be very hard for some special regions
- Performing a preliminary analysis of the area before the deployment is very important to overcome such problems

Machine Learning Approach

- Artificial Neural Networks (ANN) are an important machine learning algorithm that inspired by the brain
- ANN have many applications in the real life for prediction, classification, approximation, data processing, control
- In this study, a fast technique that is based on ANN is proposed to estimate the suitable regions
- ANN is trained by some regions features for suitability
- All other regions are defined by the ANN as suitable or unsuitable

Loading Satellite Images

- Expert loads the satellite image of the area
- That image can be searched using Google Maps Static API or loaded by hand from previously saved file

Fig. 3 : Satellite image of İzmir Gulf

Splitting the Image

- This loaded image is split into grid of square cells width edge length of $(r * \sqrt{2})$ where r is sensor coverage radius
- Expert will have a satellite image with grid cells

Fig. 4 : A sensor in center of grid

Preparing of the Training Data

- Expert must select the positive and negative cells:
  - enter 1 for the positive cells (cells to be sensed)
  - enter 0 for the negative cells (cells not to be sensed)
- Color (Red, Green, and Blue) value of all pixels of a cell is used for feature extraction
- Training data is composed of features extracted from all cells that values are set

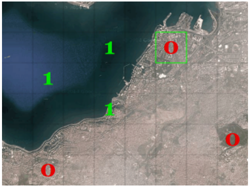

Fig. 5 : Setting positive and negative cells

Setting Expected Values

- Expert can set all the expected values of the remaining cells needed only for measuring the performance
- Expert will enter:
  - 1 to the cells expected to be positive
  - 0 to the cells expected to be negative

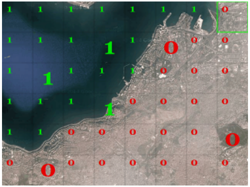

Fig. 6 : Setting expected values

Training of the ANN

- Multilayer feed forward ANN with backpropagation having:
  - 6 input neurons (Average of Red, Green, and Blue, Standard Deviations of Red, Green, and Blue):

$$x^{(i)} = \{R^{(i)}_{avg}, G^{(i)}_{avg}, B^{(i)}_{avg}, R^{(i)}_{std}, G^{(i)}_{std}, B^{(i)}_{std}\}$$
$$\text{where } 0 < i \leq n$$

  - 40 hidden layer neurons
  - 1 output neuron (positive or negative)
- For the output layer Linear Transfer Function (purelin) with threshold function:

$$output = \begin{cases} 0 & output < 0.5 \\ 1 & output \geq 0.5 \end{cases}$$

Training of the ANN

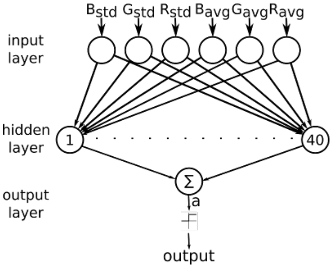

Fig. 7 : Architecture of the ANN

Results of the ANN

- Cells that are not selected for the training are estimated (positive and negative) by the ANN.
- Estimation results are compared with expected results and accuracy of the ANN is calculated

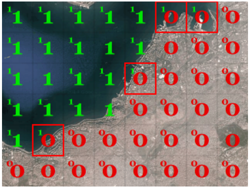

Fig. 8 : Results of the ANN (Accuracy: 91.11%)

Deployment of the Sensors

- Sensors with radius r are deployed to the center of the cells as estimated positive
- Number of sensors required to monitor the area is calculated

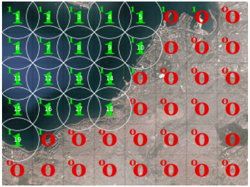

Fig. 9 : 19 sensors are deployed

Data Sets

- A set of satellite images (640x480 pixels):
  - İzmir (coordinates: 38.41.8897, 27.128677)
  - İstanbul (coordinates: 41.005294, 28.977127)
  - Karabük (coordinates: 41.211722, 32.602959)
- for five different training input sizes:
  - 2 inputs: one positive, one negative
  - 6 inputs: three positive, three negative
  - 8 inputs: four positive, four negative
  - 10 inputs: five positive, five negative
  - 16 inputs: eight positive, eight negative
- Images are split into 16x12 cells with edge of 40 pixels

Scenario 1: İzmir

- Assumed that expert wants to monitor the sea
- Positive for blue intensive areas
- Negative for brown intensive areas

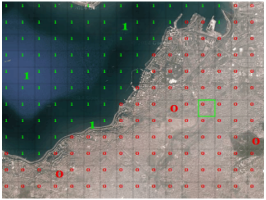

Fig. 10 : Setting inputs

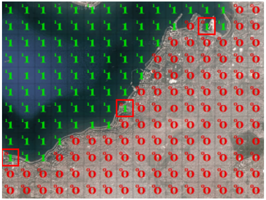

Fig. 11 : Classification result

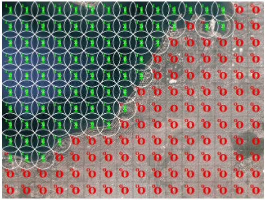

Fig. 12 : Sensor deployment

Scenario 2: İstanbul

- Assumed that expert wants to monitor the territorial areas
- Positive for brown intensive areas
- Negative for other areas

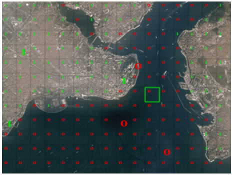

Fig. 13 : Setting inputs

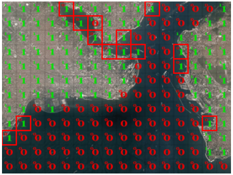

Fig. 14 : Classification result

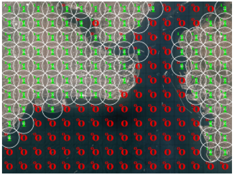

Fig. 15 : Sensor deployment

Scenario 3: Karabük

- Assumed that expert wants to monitor the forested areas
- Positive for green areas
- Negative for other areas

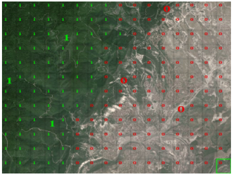

Fig. 16 : Setting inputs

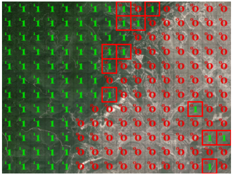

Fig. 17 : Classification result

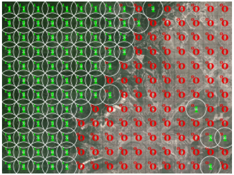

Fig. 18 : Sensor deployment

Experimental Results

- After setting training data and expected values, hence confusion matrix for the experiments can be easily generated
- Experiments are repeated for 50 times

Experimental Results

- Appropriate number of training data are required for the proper classification of the cells
- If an improper number of training data are used it may cause wrong classification, hence ineffective results
- For all experiments, based on the complexity of the image, best results have been obtained by using different number of training data:
  - 8 for İzmir
  - 16 for İstanbul
  - 16 for Karabük

Average of the Results

	2 Data	6 Data	8 Data	10 Data	16 Data
Number of sensors deployed	68.68	74.94	78.1	76.08	80.04
Accuracy	80.21%	92.46%	94.34%	93.69%	94.07%
Recall (True Positive Rate)	82.51%	95.16%	93.48%	96.09%	94.11%
Specificity (True Negative Rate)	78.96%	90.76%	95.60%	92.16%	94.04%
Precision	68.10%	86.66%	90.82%	88.70%	91.62%
F-Score	74.62%	90.71%	93.15%	92.25%	92.85%
Successful	154	171.98	173.58	170.52	165.56
Unsuccessful	38	14.02	10.42	11.48	10.44

Fig. 19 : Results for İzmir

	2 Data	6 Data	8 Data	10 Data	16 Data
Number of sensors deployed	102.62	95.8	92	96.72	91.76
Accuracy	75.04%	88.86%	90.70%	89.54%	92.11%
Recall (True Positive Rate)	69.47%	84.63%	88.00%	84.32%	88.87%
Specificity (True Negative Rate)	81.44%	93.07%	93.17%	94.84%	95.06%
Precision	81.15%	92.40%	92.19%	94.32%	94.23%
F-Score	74.86%	88.34%	90.05%	89.04%	91.47%
Successful	142.58	165.28	166.88	162.96	162.12
Unsuccessful	47.42	20.72	17.12	19.04	13.88

Fig. 20 : Results for İstanbul

	2 Data	6 Data	8 Data	10 Data	16 Data
Number of sensors deployed	82.02	79.5	79.84	81.2	73.48
Accuracy	82.05%	86.05%	87.72%	87.71%	89.84%
Recall (True Positive Rate)	75.86%	81.41%	83.23%	82.57%	89.03%
Specificity (True Negative Rate)	86.66%	89.30%	90.87%	91.42%	90.32%
Precision	80.87%	84.16%	86.47%	87.39%	84.49%
F-Score	78.28%	82.76%	84.82%	84.91%	86.70%
Successful	155.9	160.06	161.4	159.64	158.12
Unsuccessful	34.1	25.94	22.6	22.36	17.88

Fig. 21 : Results for Karabük

Conclusions

- In this paper, a fast technique to estimate the suitable regions for sensor deployment has been proposed
- Technique is tested on a set of satellite images (İzmir, İstanbul, and Karabük) with satisfactory results

Conclusions

- For an accurate classification of the regions that are suitable for deployment, an appropriate number of training data must be entered by the expert
- Expert must determine the positive and negative cells with great care for classification performance
- For the future work, it is planned to investigate the estimation of required training data number by performing a color and pattern analysis of the image and application of different classification techniques

References

1 Anna Hac, "Wireless Sensor Network designs", John Wiley and Sons, 2003

2 Genetic Algorithm - Based Sensor Deployment with Area Priority, Tahir Emre Kalaycı, Aybars Uğur, İzmir, 2011

3 Optimizing Coverage in a K-Covered and Connected Sensor Network Using Genetic Algorithms, Kasım Sinan Yıldırım, Tahir Emre Kalaycı, Aybars Uğur, 2008

4 Jennifer Yick, Biswanath Mukherjee, Dipak Ghosal, Wireless sensor network survey, Computer Networks, 2008

5 Ron Kohavi and Foster Provost, Glossary of Terms, Machine Learning, 1998, 30(2/3):271 - 274, URL: http://ai.stanford.edu/ ronnyk/glossary.html

6 `http://en.wikipedia.org/wiki/Sensor`

7 `http://en.wikipedia.org/wiki/Wireless_sensor_network`

Contact

- Enes Ateş
  enes@enesates.com
- Assoc.Prof.Dr.Aybars Uğur
  aybars.ugur@ege.edu.tr
- Asst.Prof.Dr.Tahir Emre Kalaycı
  tahir.kalayci@cbu.edu.tr