escuela técnica superior
de ingeniería informática

Label dependent evolutionary feature weighting for remote sensing data

Daniel Mateos, Jorge García and José C. Riquelme

HAIS'10 - Special Session on Hybrid ANNs: Models, Algorithms and Data

June, 23th 2010

Department of Computer Science
University of Seville

Escuela Técnica Superior
de Ingeniería Informática
Departamento de Lenguajes
y Sistemas Informáticos

Contents

Escuela Técnica Superior
de Ingeniería Informática
Departamento de Lenguajes
y Sistemas Informáticos

Introduction

- ***Remote sensing is useful for resource management, environmental monitoring, disaster response...***
- ***Land Use and Land Cover maps (LULC) are one of the main products obtained from remote sensing knowledge***
- ***LULC maps provide knowledge about the functional and morphological characteristics of the land. It is therefore, a classification problem***
- ***Several techniques have been used to develop LULC maps satisfactorily (k-NN, Naive Bayes, SVM, etc.)***
- ***Some researches have started to exploit optimization techniques for improve results***

Escuela Técnica Superior
de Ingeniería Informática
Departamento de Lenguajes
y Sistemas Informáticos

- ***Evolutionary computation is widely used to improve prediction models by weighting***
- ***There are basically three main areas of weighting application in supervised machine learning:***
  - Support vector machines optimization,
  - Artificial neural networks (training and topology),
  - Feature weighting
- ***We apply an evolutionary algorithm to search optimal weights for each feature depending on the label***
- ***The same weights are applied to all features in most existing methods***
- ***This work shows that the importance of each feature can depend on the class to predict***

Escuela Técnica Superior
de Ingeniería Informática
Departamento de Lenguajes
y Sistemas Informáticos

Data description and preprocess

- ***The data for this study belongs to the north of Galizia (Spain)***
- ***The data was obtained from fusion of sensors: LIDAR (Light Detection and Ranging) + orthophotograph***
- ***LIDAR is a laser-based sensor technology to determine distance to an object or surface***
- ***Orthophotograph is an aerial photograph geometrically corrected such that the scale is uniform***

Escuela Técnica Superior
de Ingeniería Informática
Departamento de Lenguajes
y Sistemas Informáticos

- ***Every instance has 61 basic statistics from:***
  - LIDAR data: height and intensity
  - Image data: red band, green band and blue band
- ***Every instance has 5 classes: road, farming land, middle vegetation, high vegetation and buildings***
- ***Three different filters are executed:***
  1. Each missing value is replaced with the average value of the corresponding feature
  2. Data is standardized
  3. Correlation Feature Selection method (CFS) is applied in order to reduce the search space (18 features are selected)

Escuela Técnica Superior
de Ingeniería Informática
Departamento de Lenguajes
y Sistemas Informáticos

Algorithm

- ***The goal of the proposed evolutionary algorithm is to find an optimal set of weights to improve the classification process***
- ***This set of weights depends on each label and it makes a linear space transformation***
- ***After the evolutionary execution the classification process is performed in two steps:***
  - The weights are applied to the training instances according to its label
  - Given a test instance, the label of the transformed nearest neighbour is chosen

Algorithm

- ***Build the initial population of individuals***
- ***Evaluate the fitness of each individual and save the best individual***
- ***Repeat until termination***
  - Select several individuals for reproduction according to a criterion
  - Create new individuals through crossover and mutation operations
  - Evaluate the fitness of new individuals and save the best individual
  - Replace the population with new individuals

Algorithm – Initial population

An individual is a matrix which represents the weights per label for every feature

$$\begin{bmatrix} w_{11} & \cdots & w_{1f} \\ \vdots & \ddots & \vdots \\ w_{b1} & \cdots & w_{bf} \end{bmatrix}$$

where there are b rows (number of different labels), and f columns (number of features)

Algorithm – Crossover and mutation

- ***Given two individuals selected by roulette-whel method, the i^{th} row of one is crossed with the i^{th} row of the other one***
- ***Two rows are crossed by two ways:***
  - **Uniform crossover**: selection of a weight from one parent at random
  - **BLX-α crossover** : If w_1 and w_2 are the i^{th} weight from each parent, the new weight is a real number randomly selected in the interval $[W_{min}-I\alpha, W_{max}+I\alpha]$, where:

α=positive real number,

$W_{max}=\max(w_1 , w_2)$,

$W_{min}=\min(w_1 , w_2)$,

$I=W_{max}-W_{min}$

Escuela Técnica Superior
de Ingeniería Informática
Departamento de Lenguajes
y Sistemas Informáticos

- ***The mutation operator increases or decreases the value of a weight according to a probability p. The increase (or decrease) is a random value Δ that satisfies:***

$$\Delta = \frac{r}{10^z},\ where:$$

$$r \in \mathbb{R} : [0 \leq r \leq 1]\ and$$
$$z \in \mathbb{Z} : [0 \leq z \leq n]$$

Escuela Técnica Superior
de Ingeniería Informática
Departamento de Lenguajes
y Sistemas Informáticos

Algorithm – Fitness function

W is the matrix $\begin{bmatrix} w_{11} & \cdots & w_{1f} \\ \vdots & \ddots & \vdots \\ w_{b1} & \cdots & w_{bf} \end{bmatrix}$

fitness=0
for $i = 1$ to m **do**
    We divide P into n bags: $B_1, ..., B_n$
    for all bag B_k **do**
        We apply the W transformation to every point from the remaining $n-1$ bags, obtaining the set of points P'
        for all point p_i in B_k **do**
            for all label $l \in \{1..b\}$ **do**
                We construct the tranformed point p_i^l so that $p_{ij}^l = w_{lj} * p_{ij}$
                We calculate d_l = minimum distance from p_i^l to the points of P'
                We apply the W transformation to p_i according to its label, and we add it to P'
            end for
            We calculate the minimum from the distances d_l. Let $h \in \{1..b\}$, the label of the point of P' which makes d_l.
            if the label of $p_i \neq h$ **then**
                $fitness = fitness + 1$
            end if
        end for
    end for
end for

Escuela Técnica Superior
de Ingeniería Informática
Departamento de Lenguajes
y Sistemas Informáticos

Results

- ***Evolutionary algorithm setup:***
  - Population of 20 individuals
  - 100 generations
  - 20% of mutation probability
- ***Testing:***
  - Competitors: Naive Bayes, SMO, Nearest Neighbour, Multilayer Perceptron
  - Stratified 3 x 10-fold cross-validation
  - Friendman test, in which the null hypothesis is that all classifiers have the same performance

Escuela Técnica Superior
de Ingeniería Informática
Departamento de Lenguajes
y Sistemas Informáticos

Results

Algorithm	**Error**
Naive Bayes	0.15
SMO	0.14
Nearest Neighbour	0.13
Neural Network	0.10
Weighted-Nearest Neighbour	0.10

Table 1: **Averaged error rate for each studied algorithm**

Escuela Técnica Superior
de Ingeniería Informática
Departamento de Lenguajes
y Sistemas Informáticos

Results

Class	Features			
Road	MINSNDVI	PEC	IMAX	HCV
Farming Land	IMEAN	IGMEAN	HMAX	HCV
Middle Vegetation	HSTD	IGKURT	MINSNDVI	IGVAR
High Vegetation	IMAX	IRVAR	IGVAR	IGMEAN
Buildings	IGKURT	PCT32	EMP	MINSNDVI

H*: height statistic, I*: Intensity statistic, IG*: Intensity green band stat., IR*: Intensity red band stat., *SNDVI: Simulated Normalized Difference Vegetation Index stat., PEC: Penetration coef., PCT32: Percentage third or later returns over second returns, EMP: Empty pixels that surrounds the current pixel

Table 1: **Most important features according to its weight for the study zone.**

Escuela Técnica Superior
de Ingeniería Informática
Departamento de Lenguajes
y Sistemas Informáticos

Conclusions

- ***We have presented a simple method to transform the feature space***
- ***Different weights are assigned to every feature depending on each class***
- ***The results showed an improvement of the 3% on the Nearest Neighbour, and resemble the results of the best competitor***
- ***We provide valuable information about the importance of each feature in order to distinguish among the different classes***
- ***The accuracy improvement and the algorithm definition as an independent preprocessing method, is the aim of future work***

Escuela Técnica Superior
de Ingeniería Informática
Departamento de Lenguajes
y Sistemas Informáticos

References

[1] Erdody, T., Moskal, L.: Fusion of LIDAR and imagery for estimating forest canopy fuels. Remote Sensing of Environment (to appear) (2010)

[2] Atkinson, P.M.: Spatially weighted supervised classification for remote sensing. International Journal of Applied Earth Observation and Geoinformation **5**(4) (2004) 277 – 291

[3] Bork, E.W., Su, J.G.: Integrating lidar data and multispectral imagery for enhanced classification of rangeland vegetation: A meta analysis. Remote Sensing of Environment **111**(1) (2007) 11 – 24

[4] Mazzoni, D., Garay, M.J., Davies, R., Nelson, D.: An operational misr pixel classifier using support vector machines. Remote Sensing of Environment **107**(1-2) (2007) 149–158

[5] Shao, G., Wu, J.: On the accuracy of landscape pattern analysis using remote sensing data. Landscape Ecology (23) (2008) 505–511

[6] Tomppo, E.O., Gagliano, C., Natale, F.D., Katila, M., McRoberts, R.E.: Predicting categorical forest variables using an improved k-nearest neighbour estimator and landsat imagery. Remote Sensing of Environment (113) (2009) 500–517

[7] Howley, T., Madden, M.G.: The genetic kernel support vector machine: Description and evaluation. Artificial Intelligence Review **24** (2005) 379–395

[8] Hervás-Martínez, C., Martínez-Estudillo, F., Carbonero-Ruz, M.: Multilogistic regression by means of evolutionary product-unit neural networks. Neural Networks **21**(7) (2008) 951 – 961

[9] Komosinski, M., Krawiec, K.: Evolutionary weighting of image features for diagnosing of CNS tumors. Artificial Intelligence in Medicine **19**(1) (2000) 25–38

escuela técnica superior
de ingeniería informática

Thanks for your patience !

{mateos, jgarcia, riquelme}@lsi.us.es

Department of Computer Science
University of Seville