Overview on spatial classification methods with emphasis on spatial autocorrelation

KIRSI VIRRANTAUS

GIS-E4020

14.2.2019

1. Back to data mining

During this course you have learned some most used data mining methods and their spatial versions

Association rule

Clustering

SOM

GWR

Trajectory data mining

In this lecture we go back to one of the core methods - **classification**

Spatial classification

In classification the problem is that we want to create a function that can be used in **predicting** the dependent value by using data on the independent variables, example

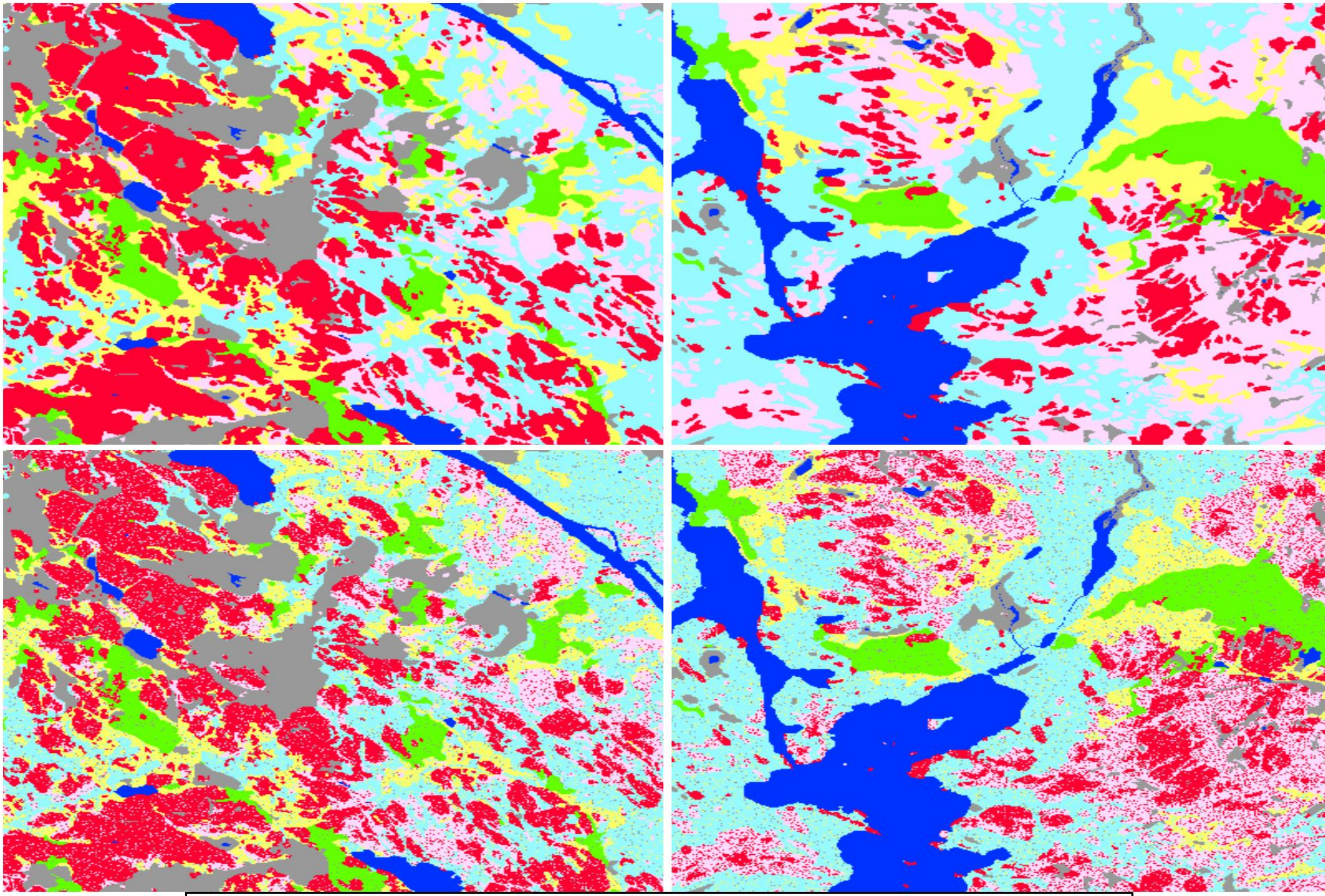
- classification in remote sensing: land use classes

The function is found by **learning**; it means that we have data sets that include both dependent and independent variable data

Spatial classification has a couple of special problems

- **“Salt and pepper”** problem means that classification focuses only at one location at a time, spatial autocorrelation is not taken into account properly and the result is scattered
- The other problem is related to **classification accuracy**

6.1



Salt and pepper

Classification methods

Shekhar et al. (2002) mention three classification methods that have spatial extension

Markov Random Field Bayes classifier

Spatial Autoregression method

Geographically Weighted Regression

Markov Random Field Bayes classifier (MRF classifier)

So-called naïve Bayes classifier

Can be described as a tree; classes have probabilities

The disadvantage is that the regular Bayeds classifier method is local (salt and pepper)

Using MRF classifier means that the **neighbourhood pixel probabilities** are taken into account

Markov property: “memoryless” property of a stochastic process

- Markov chain in 2d: 4 neighbourhood

In Bayes the a posteriori probability is based on the knowledge on a priori probabilities; the probabilities of the neighbouring pixels effect to the probability of the pixel in focus

Conditional Autoregressive Process

Marcov Random Field Autoregressive Process is on special type of conditional autoregressive processes (CAR)

http://www.statsref.com/HTML/index.html?car_models.html

Idea: the probability of values estimated at any given location are conditional on the level of neighboring values

Logistic spatial autoregressive model (SAR)

In spatial autoregressive model the dependent variable is calculated by solving

$$y = \rho W y + X\beta + \epsilon.$$

Logistic regression is a version of regression model that can be used for binary dependent variables

The models presents the **dependencies between the binary dependent variable and one or more independent variables** (nominal, ordinal, interval or ratio-level variables)

SAR models – spatial lag in general

http://www.statsref.com/HTML/index.html?sar_models.html

A pure SAR model consists of a **lagged version of regression model**

Idea: The dependent variable is dependent on the values of neighbouring locations

$$X = \rho W X + \varepsilon$$

$$X = (I - \rho W)^{-1} \varepsilon$$

W-matrix contains adjacency information
Rho stands for the strength of autocorrelation
Epsilon is normally distributed random value
X is the dependent variable

Geographically Weighted Regression GWR

The limitation of SAR is that it does not take geographical heterogeneity into consideration

Parameter estimates and error are same throughout the space

In Geographically weighted regression the heterogeneity is modeled

$$y = X\beta(s) + \epsilon(s),$$

Where **parameter estimates and error are spatially varying**

2. About classification methods in general

Aspatial classification methods that do not take autocorrelation or heterogeneity in to account can of course be used in local mode

Popular method are for example:

Nearest neighbor

Random forest

Neural networks

Support vector machines

Decision trees

3. Decision tree

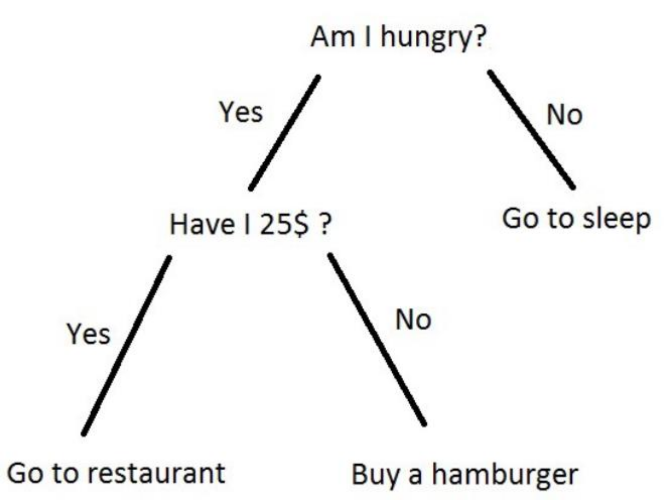
Decision tree (päätöspuu) is a decision support tool that uses a tree-like model of decisions and their possible consequences

Decision trees are used in

operations research in decision

analysis

- Simple example



Decision tree is also used in machine learning, especially in classification tasks in order to be used in prediction

Decision tree in classification

In decision tree classification **a decision tree** is used to go from observations of an item to conclusions about the item´s target value (class)

In the tree leaves represent **class labels** and the tree structure represents the logic according to which the label is decided; in nodes the conditions are presented

The **learning task** is to create a tree that best characterizes the relationship between the explanatory features and the ground truth class labels in the data

In spatial classification we aim to minimize both the **classification errors** and **salt and pepper noise**

Entropy and information gain

In the training of classifier the goal is to cause as pure classes as possible (items in one class is below and in another class in above the threshold)

Entropy (entropia = haje, epäjärjestyksen määrä) is the measure of this impurity of class distributions

Information gain (informaatiohyöty) is the decrease of entropy after a split

In the spatial decision tree there are a couple of other measures (neighbourhood split autocorrelation ratio and spatial information gain)

In training the tree these measures are used (here we do not go to details in this) – the algorithm is describes in detail on the p 68 in (Jiang&Shekhar, 2017)

(Jiang & Shekhar, 2017)

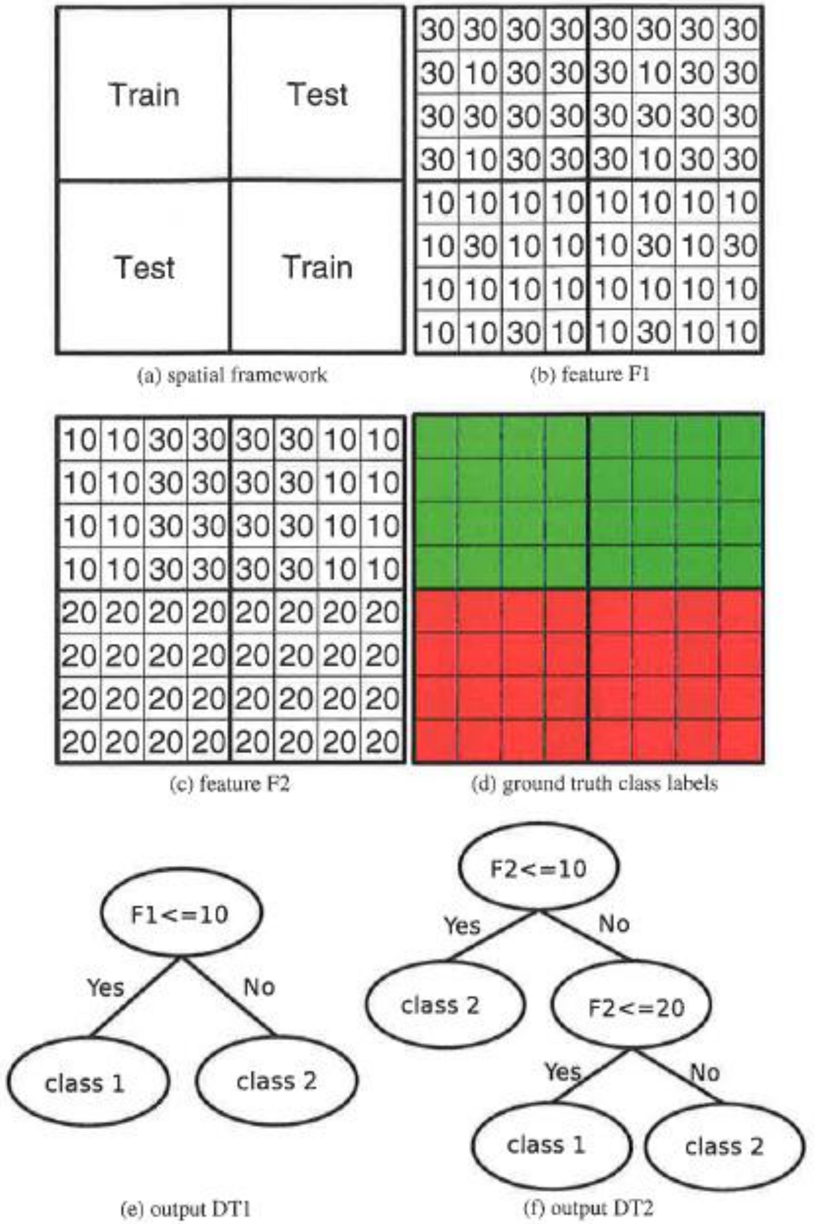


Fig. 4.4 Example of problem input and output with *red* for class 1 and *green* for class 2 (best viewed in color)

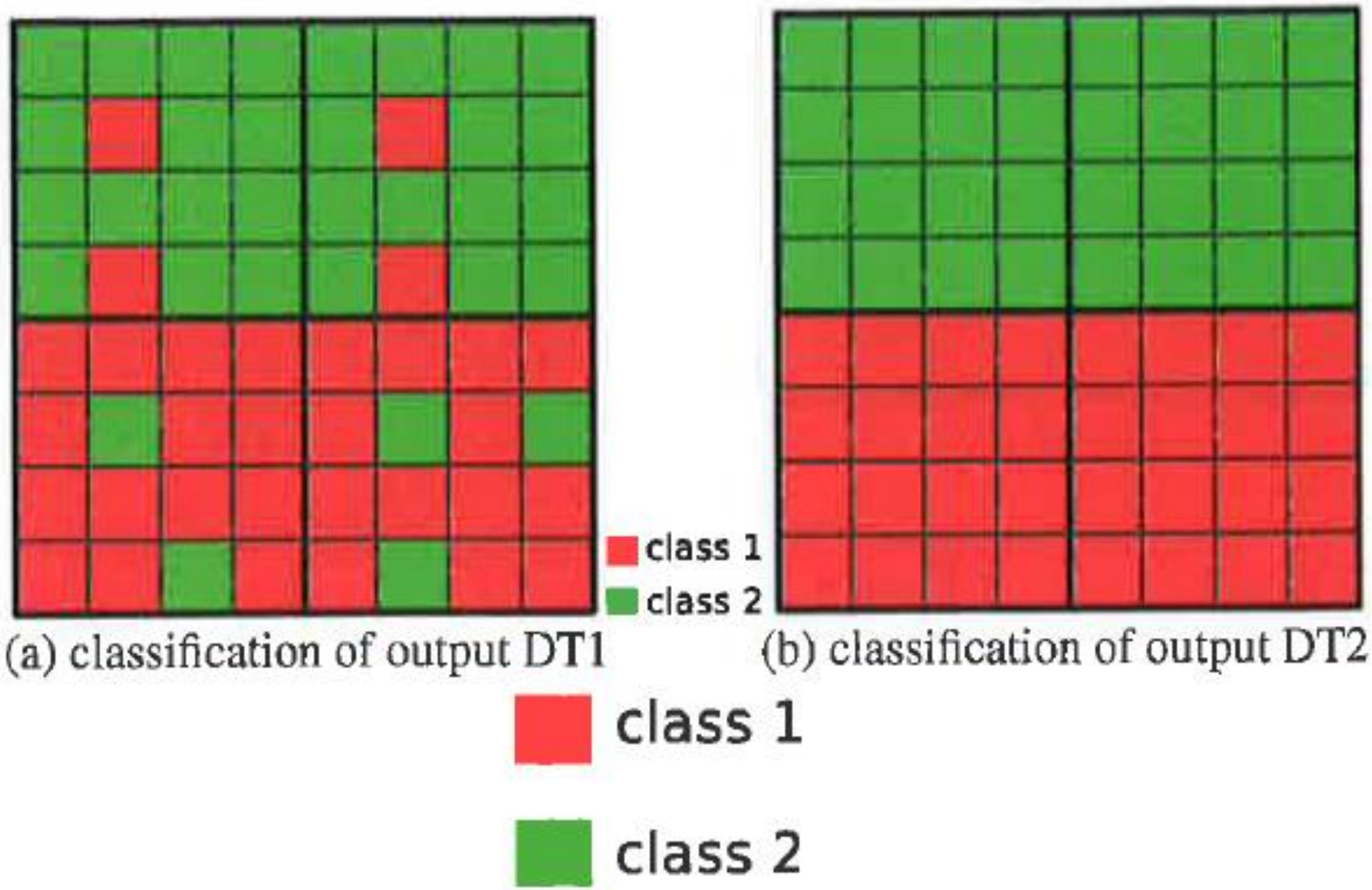


Fig. 4.5 Classification results of decision tree classifiers from Fig. 4.4e, f

Jiang&Shekhar, 2017)

Spatial decision tree

In spatial decision tree the data is spatial

Spatial characteristics can be presented

- By **spatial predicates** (distance, direction) in the conditions
- By using **topological relationships** (adjacency) – for example in grid structure

Koperski et al. have published an early paper on using spatial predicates in spatial decision trees

The following method by Jiang&Shekhar uses grid and focal neighbourhood method

Focal-test –based spatial decision tree

In this presentation the method developed by Zhe Jiang and Sashi Shekhar is presented (2017)

It offers a relatively simple method for developing decision tree classifier into spatial version

The classification is based on both **local information** and **focal information**

Focal test –refers to the method: the neighbourhood area (for example 3x3 in grid) is considered

A so-called **gamma index** is created that shows the amount of autocorrelation in the neighbourhood

Focal neighbourhood concept

Spatial neighbourhood concept is based on 3x3 window

Adjacency relationship is 4-adjacency and presented as w-matrix

Spatial autocorrelation is described by a statistic that measure the degree of dependency among attribute values of neighbouring grid cells

The measure is called Gamma index

i, j are any pair of grid cells

a is spatial similarity and b is class similarity

w matrix represents the spatial similarity (adjacency)

delta represents the attribute similarity (1,0)

Gamma index

$$\Gamma_i = \frac{\sum_j W_{i,j} I_i I_j}{\sum_j W_{i,j}},$$

Local and focal test

Local test:

- tests the value of a feature (f) in location against a threshold; threshold is marked by delta δ; **test f =< δ**

Local indicator:

- show the result of local test in each location I = 1 or 0

Focal autocorrelation statistic, focal indicator:

- **Gamma index** Γ summarizes the values of local indicators in the neighbourhood; if index < 0 it indicates that this location is potentially salt and pepper noise

Focal test: Gamma index value

Γ <0 potential salt and pepper

$$\Gamma_i = \frac{\sum_j W_{i,j} I_i I_j}{\sum_j W_{i,j}},$$

Decisions in the decision tree

Classification decisions are made in the nodes of decision tree

Local and focal test results are combined

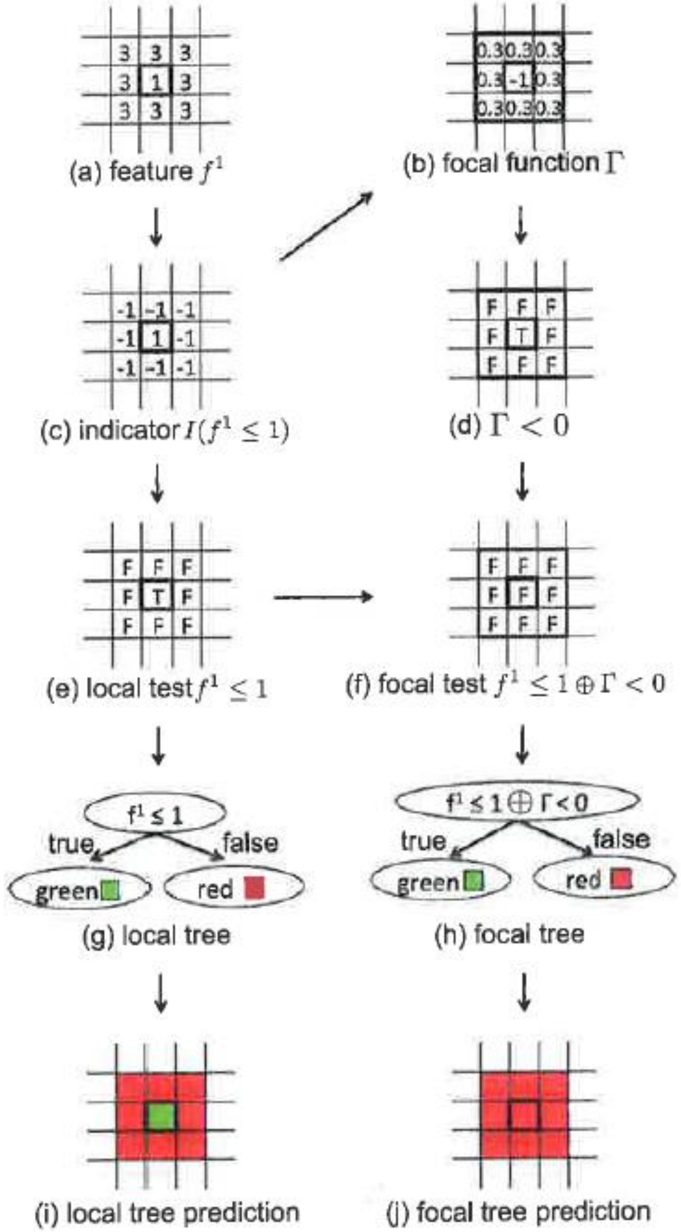
XOR means that **one and only one is true**

- If Local test is true and Focal is not; if Focal test is true but Local test is not

Example on XOR reasoning

Local test	Focal test	Result
belongs to the class	potential salt and pepper	0
belongs to the class	not salt and pepper	1
does not belong to the class	potential salt and pepper	0
does not belong to the class	not salt and pepper	1

Fig. 5.4 Comparison of a local test versus a focal test, a local-test-based decision tree versus a focal-test-based spatial decision tree. ("T" is "true"; "F" is "false")



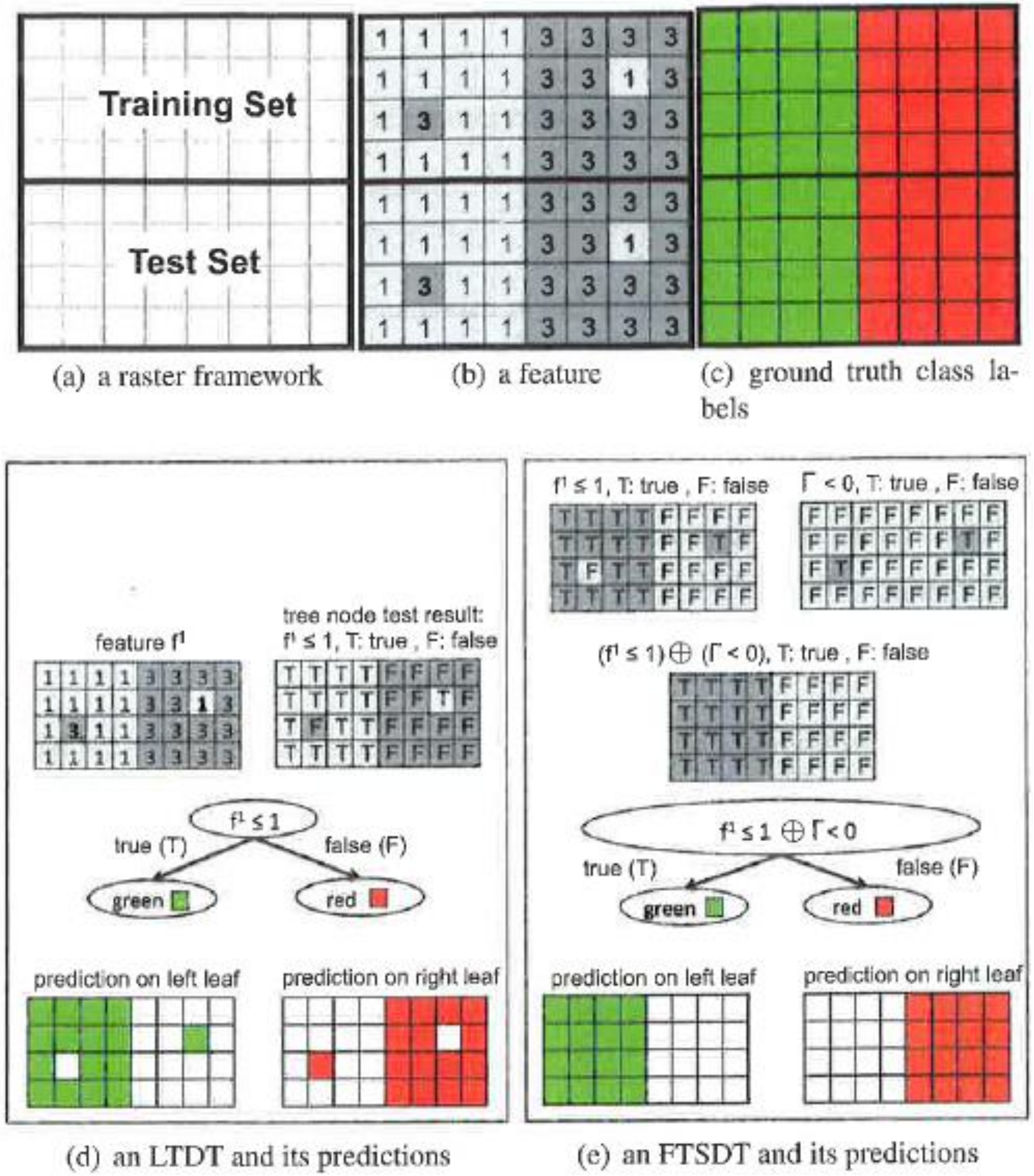


Fig. 5.5 Illustrative problem example (best viewed in color)

Literature

Jian,Z., Shekhar,S., Spatial Big Data Science, Chapters 2 and 4. 2017.

Shekhar,S., Schrater,P., Vatsavai,R., Wu,W., Chawla,S., Spatial Contextual Classification and Prediction Models for Mining Geospatial Data, IEE, 2002.

Shekhar,S., Vatsavai,R., Chawla,S., Chapter 6 (Spatial Classification and Prediction Models for Geospatial Data Mining) in book "Geographic Data Mining and Knowledge Discovery" by Miller,H., and Han,J., 2009.