

Satellite image classification using granular neural networks

D. STATHAKIS*^{†‡} and A. VASILAKOS[†]

[†]Department of Planning and Regional Development, University of Thessaly,
Pedion Areos 38334, Greece

[‡]EC Joint Research Centre, IPSC, MARS-FOOD, 21020 (VA), Italy

(Received 19 May 2005; in final form 7 January 2006)

The increased synergy between neural networks (NN) and fuzzy sets has led to the introduction of granular neural networks (GNNs) that operate on granules of information, rather than information itself. The fact that processing is done on a conceptual rather than on a numerical level, combined with the representation of granules using linguistic terms, results in increased interpretability. This is the actual benefit, and not increased accuracy, gained by GNNs. The constraints used to implement the GNN are such that accuracy degradation should not be surprising. Having said that, it is well known that simple structured NNs tend to be less prone to over-fitting the training data set, maintaining the ability to generalize and more accurately classify previously unseen data. Standard NNs are frequently found to be accurate but difficult to explain, hence they are often associated with the black box syndrome. Because in GNNs the operation is carried out at a conceptual level, the components have unambiguous meaning, revealing how classification decisions are formed. In this paper, the interpretability of GNNs is exploited using a satellite image classification problem. We examine how land use classification using both spectral and non-spectral information is expressed in GNN terms. One further contribution of this paper is the use of specific symbolization of the network components to easily establish causality relationships.

1. Introduction

During the past few years, satellite image classification to produce land use or land cover maps has shifted from finding the right data to finding a method able to cope with the plethora of available data. Today many different satellites orbiting the earth provide data with increased spectral, spatial and temporal resolution. In addition, topographic data are now more easily obtained. Digital Elevation Models (DEMs) are available at resolutions ranging from approximately 1 km (*Shuttle Radar Topography Mission*, SRTM 2000) to a few tens of centimetres (*Light Detection And Ranging*, LIDAR). Two other very useful sources of information, namely slope and aspect, are commonly derived from the DEM. Consequently, to study a phenomenon such as land use, we now have a choice of suitable spectral, spatial and temporal resolution and topographic data.

To combine all this information in a classification system many have resorted to the neural networks (NNs) paradigm. The use of neural networks is promising as it

*Corresponding author. Email: dstath@uth.gr

offers at least comparable accuracy with respect to conventional methods and at the same time the ability to work with data not fully conforming to statistical distributions. The down side is that the structure of neural networks has been difficult to deduct from the problem at hand. To form the actual structure of the neural network, decisions are made based on experience and rules of thumb (Kanellopoulos and Wilkinson 1997). It is difficult and time-consuming to understand which factors and to what extent influence the output (Lippmann 1987, Ito and Omatu 1997). Usually the structure is labyrinthine, full of connections with many different numeric weights that are hard to explain. For this reason neural networks have often been accused of being black box techniques. Needless to say that the application of NNs in any scientific field depends on the confidence we have in understanding the process taking place.

Recently, the scientific field of neural networks has merged with that of fuzzy sets to form neural networks based on information granules, called granular neural networks (GNNs; Pedrycz and Vukovich 2001). To form the granules, fuzzy set theory is deployed (Zadeh 1965) in an effort to shift our concern to the meaning of the information rather than its measure (Zadeh 1975). Each NN component has an exact meaning that can be stated in linguistic terms and vice versa. Land cover type is clearly a human concept that is a matter of degree (Zadeh 1999). Methods where the processing is done on a numerical level where arbitrary thresholds are chosen are clearly limited. Processing on a conceptual level can be achieved by merging membership functions of different factors expressed as linguistic terms (Bortolan 1998, Pedrycz and Vasilakos 1999).

In the remote sensing data context, it might prove more efficient to treat spectral and non-spectral information as linguistic rather than numerical variables. Infrared for example, can be a linguistic variable where, with respect to reflectance, we have low, medium and high and not the actual numeric values stored for each pixel. Linguistic variables provide the means to establish approximate reasoning of over-complex phenomena (Zadeh 1975). If we accept that high precision is incompatible with high complexity (Zadeh 1975), our method would be more suited to operating on information granules rather than on vast amounts of numerical values.

The synthesis of fuzzy sets and neural networks yields another very important benefit, interpretability. Granular neural networks compete with other methods, including standard neural networks, in interpretability of results rather than accuracy (Bortolan and Pedrycz 2002, Pedrycz and Reformat 2005, Nauck 2003). The implementation of GNNs has been based on both feed forward NNs and Kohonen's self organizing maps (Lin *et al.* 2000). Optimization via genetic algorithms (GAs) has also been proposed (Mitra *et al.* 2002, Shapiro 2002, Pedrycz and Reformat 2005).

Qiu's and Jensen's (2004) means of opening the neural network black box was a granular neural network based on a Kohonen's self-organizing map. They were able to interpret the structure of the network, but with slightly complicated rules since fuzzy sets are described with actual numeric parameters rather than linguistic tags. Their implementation outperformed a standard back-propagation trained network and a maximum likelihood classifier yielding overall acceptable level of accuracy. Dixon (2004) deployed a GNN to investigate its applicability in solving spatial problems, predicting ground water vulnerability in specific. Interestingly, she conducted a sensitivity analysis to investigate how different parameters, such as

shape selected to approximate membership and number of fuzzy sets per variable, affect classification accuracy.

The adaptive neural fuzzy inference system (ANFIS) has been also used to classify remote sensing data into 10 land use classes (Benediktsson and Benediktsson 1999). Topographic data, namely elevation, slope and aspect, were used in addition to spectral information, a four-channel Landsat MSS image in particular. Performance comparison to other methods showed that the neuro-fuzzy method yields superior results for certain configurations. The neuro-fuzzy model has been found to be more suitable for classifying a mixed-composition natural environment using very high-resolution images of IKONOS satellite (Han *et al.* 2002). This model performed better compared with a back-propagation neural network and a maximum likelihood classifier. The neuro-fuzzy method was more efficient in classifying natural environment classes, such as forest and vegetation, than human made structures such as roads.

The novelty of the present paper is the application of the GNN method to the remote sensing imagery classification context. In addition, it is shown that GNN is a suitable framework for the concurrent use of spectral and non-spectral information in the classification. Another fundamental contribution of our approach is the establishment of the conceptual link between neural network components and linguistic rules in the context of remote sensing classification. The use of specific symbolization further contributes to clarify this bidirectional link. We proceed by presenting the GNN architecture in §2. A description of its application in land use classification is given in §3. The discussion on the classification results follows in §4 and then conclusions are presented.

2. GNN architecture

The concept of granules is central. Granules are information clusters based on a model such as the fuzzy sets, rough sets and shadowed sets (Pedrycz and Vasilakos 1999). In the present work, fuzzy sets are selected as the information granulation means. In the remote sensing context, the pixel's digital value in each information band is a data point. The data points are clustered together to form granules based on similarity. The membership function is used to model similarity.

In addition, the use of fuzzy sets enables us to extract knowledge from the NN skeleton. We are able to convert the fuzzy IF ... THEN rules to neural network skeletons and vice versa. Let us say for example that we have the following simple rule, where fuzzy sets corresponding to linguistic labels are marked with letters in italics.

RULE A: If rainfall is *heavy* and altitude is *low* then yield is *good*.

Rainfall and altitude are our input variables. Yield is the output variable. The corresponding network skeleton is shown in figure 1. We need at least two fuzzy sets per input and output class for our example to make sense. Let us suppose that we have heavy and light fuzzy sets for rainfall and low and high for altitude. Yield is categorized as good or bad. Suppose that the rule for bad yield is the following:

RULE B: If rain is *light* and altitude is *high* then yield is *bad*.

The rule set (Zadeh 1997) is a collection of linguistic rules. The network skeleton of both rules is shown in figure 2. Fuzzy set theory is an important aid in building transparent NNs with respect to the spatial data context, by conforming to the necessity of abandoning precision and becoming more tolerant in approximating nature (Zadeh 1975). It is also interesting that we can state non-numeric perceptions

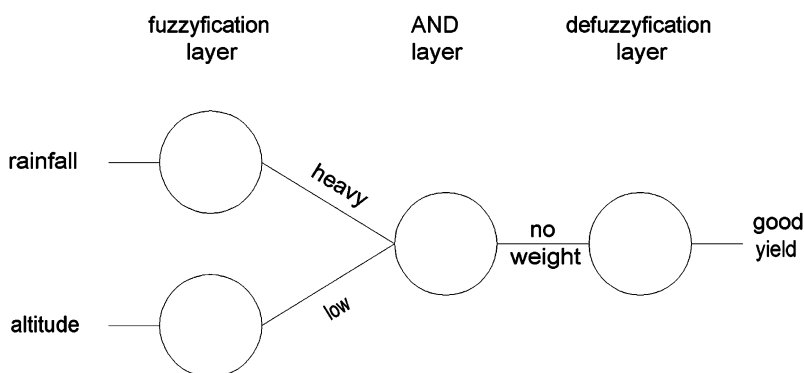


Figure 1. Rule A is converted to a neural network skeleton in exact terms. The weights of lines are fuzzy sets rather than the usual numerical values.

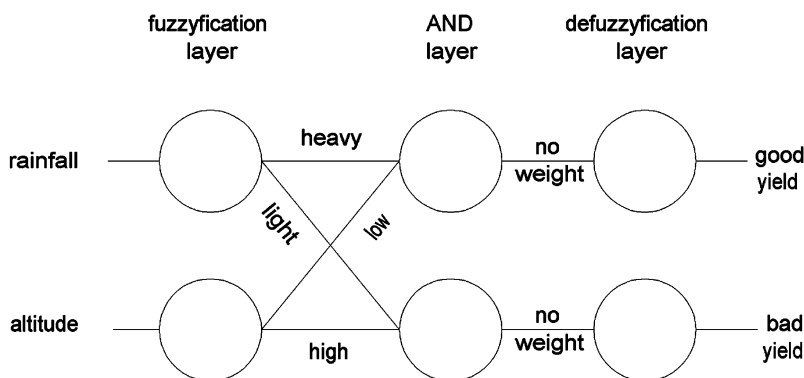


Figure 2. Complete conversion of the rule set, consisting of rule A and rule B, into a neural network structure.

together with numeric ones. One example could be the following hypothetical rule (not implemented in the present work) in a food security prediction system:

RULE C: If rainfall is *high* and political_situation is *not_stable* then status is *aid_needed*.

The specific GNN architecture used is based on the NEFCLASS-J software (Nauck 1999, 2003). The following characteristics of the system are organized and presented in neural network terms to facilitate comparisons with standard neural networks trained by the back-propagation algorithm (Rumelhart *et al.* 1987) as well as other types of networks.

The architecture of the system can be realized as a feed-forward neural network with just one hidden layer since only the AND fuzzy rule is implemented. This layer consists of AND neurons thus we could call it the AND layer. It should be noted here that an OR layer has also been proposed (Pedrycz and Reformat 2005) as an additional layer, consisting of the OR neurons, aimed at the aggregation of the rules having the same conclusion. It is possible that pruning reduces the need for the OR layer, but we are not sure at this point if it totally eliminates the need for it. It is also important to keep in mind that, according to the Kolmogorov theorem (Atkinson

and Tatnall 1997), an NN with just one hidden layer is able to represent any function regardless of complexity. The fuzzyfication (in) and defuzzyfication (out) layers are also present.

The number of nodes in the hidden layer is equal to the number of rules. Sigmoid activation function is used for the hidden layer and maximum activation function for the output layer. This means that, where we have more than one rule leading to the same output class, the one producing the maximum activation is selected. The final class is chosen using the winner takes all rule. In other words, the network operates with fuzzy sets but is forced to produce crisp output at the end.

The network could be potentially fully connected but usually in practice it is not. Inputs to hidden layer connections are created depending on the composition in terms of linguistic tags of each rule. The link is realized only if the input variable is present in this rule. From hidden to output layer a connection is created if the rule contributes to the output class (Nauck 2003). As shown in figure 2, the weights of the synapses are fuzzy sets rather than the classic numeric values. One important aspect is that, to maintain interpretability, hidden to output links have no weights (Nauck and Kruse 1999). Some links share weights to enforce the fact that each linguistic term must have only one meaning.

A backpropagation-like algorithm is used to perform supervised learning. Error is propagated backwards, but a simpler technique than gradient descent is used. Gradient descent methods cannot be applied directly to a fuzzy system, because the functions used in the inference process are usually not differentiable (Nauck 2003). The training method used directly alters the fuzzy set parameters during the learning phase (Wang and Mendel 1992). The magnitude of correction applied is proportional to the output error produced. More importantly, we have the option of pruning the network, so that the same level of accuracy is achieved by less input nodes and links. In effect the architecture is actually changed by reducing the number of nodes in the hidden layer and also the connections. It should be noted that the overall process is governed by the notion that interpretability is preferred over accuracy. The general study flow is not different from that followed in standard neural network classification. It includes normalization of input, splitting available reference samples into two partitions, the training and testing sets. Instead of the standard NN, the GNN is used. The trained network is characterized by the fact that it can be described in terms of linguistic rules.

3. Interpreting remote sensing data classification for land use

To validate the use of GNN with remotely sensed data we have implemented a model based on a satellite scene of central Greece and a digital elevation model derived from a medium scale (1:50000) topographic map. The original 20 m resolution contours were digitized and the DEM was resampled to match the resolution of the satellite image (23.5 m). We used a granular neural network implemented in the NEFCLASS-J software to classify spectral and non-spectral information into land cover classes. The spectral information consisted of green (0.52–0.59 μm), red (0.62–0.68 μm), near-infrared (NIR, 0.77–0.86 μm) and middle-infrared (MIR, 1.55–1.70 μm) bands of a satellite image acquired during July 1998 by the Indian Remote Sensing (IRS1D) LISS III sensor (figure 3). To increase the dimensionality of information, aiming at better distinguishing between categories, we also used non-spectral information, namely slope, aspect and elevation. Slope and aspect are derivatives of elevation. Percentage slope was computed using a



Figure 3. Satellite image used. Spectral values are stretched here for clarity.

moving 3×3 kernel (Burrough and McDonell 1998):

$$\text{slope} = \sqrt{\frac{(z_{NE} + 2z_E + z_{SE}) - (z_{SW} + 2z_S + z_{NW})}{8X_{\text{cellsize}}}} + \sqrt{\frac{(z_{SW} + 2z_S + z_{SE}) - (z_{NW} + 2z_N + z_{NE})}{8Y_{\text{cellsize}}}}$$

where z_i is the elevation of cell i in the kernel where i corresponds to orientation with respect to the centre of the kernel (i.e. Z_{NE} is the north-east neighbour of the central pixel and so on). X_{cellsize} and Y_{cellsize} are the size of the pixel. In our case both variables equal 23.5 m. The aspect compares the eight cells forming the neighbours of each pixel and selects the one with the lowest elevation. The aspect is then the azimuth value according to direction to the lowest elevation (downslope). If the direction is north, then the aspect is 0; if the direction is north-east then aspect is 45, and so on for all possible directions clockwise.

The original input vector is thus:

[red, green, NIR, MIR, slope, aspect, elevation]

Whereas the output vector, containing the desired land use categories is:

[urban dense, urban sparse, forest, agricultural, water, soil/sparse vegetation]

By pruning the classifier, the dimensionality of the input vector is reduced without significant loss in accuracy. The stopping rule for pruning is to have at least one rule per output class and at the same time remove a rule only if the action does not result in performance degradation. It has been found that using half of the available input dimensions can sometimes result to only 1% loss in classification accuracy (Benediktsson and Sveinsson 1997).

MIR band and slope are removed and the pruned input vector takes the following form:

[red, green, NIR, aspect, elevation]

Figure 4 is the conceptual diagram of the pruning process for our case study.

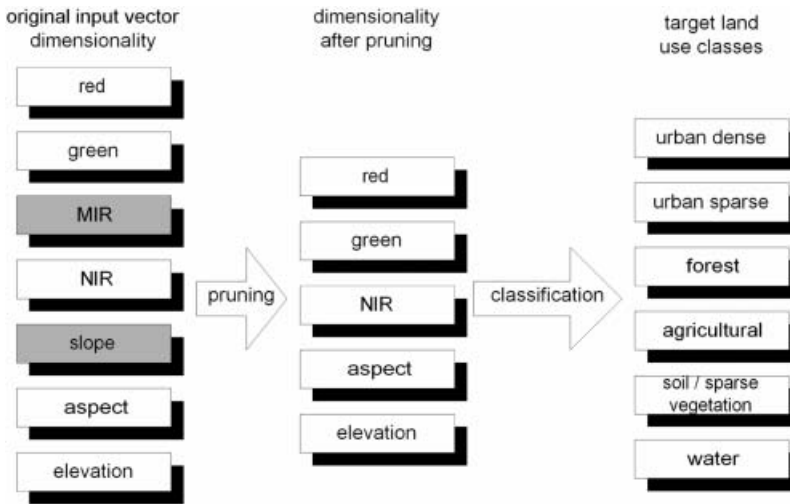


Figure 4. Dimensionality of input vector is reduced without significant loss in classification accuracy, but with increased contribution to interpretability. The output vector, target land use classes, remain of course constant.

To train and test the system, we used labelled input vectors. The labels came from field visits where location was recorded using a global positioning system (GPS), and land use was determined *in situ*. The target number of samples was 50 per class so that a confusion matrix (Congalton and Green 1998) could be constructed. The total achieved number of samples was 266, comprising 56, 32, 29, 64, 50 and 35 per urban dense, urban sparse, forest, agricultural, soil/sparse vegetation and water accordingly. To visit the field essentially means that any form of random sampling, systematic or stratified, cannot be adopted since it is not feasible to gain access to many sites due to natural and legal barriers. Given that it is often impractical to follow such schemas, we tried to balance statistical requirements with practicalities (Foody 2002). Bearing that in mind, our target was to collect around 50 samples per class while at the same time trying to keep the samples independent, to avoid spatial autocorrelation effects, and as representative as possible.

All input dimensions were normalized to the range $[0,1]$ of real numbers, thus the reflectance value of 100 in a 7 bit image channel becomes 0.781, for example. Output classes, i.e. land use types, were binary coded so that $\{1,0,0,0,0,0\}$ represents urban dense, etc. Alternative output coding schemas (Ersoy and Hong 1990) were not investigated here. Half of the available labeled vectors were used for training and the other half for testing. The bell shape was selected as a fuzzy function approximation. Three fuzzy sets per class were defined as low, medium and high and their domains were always the $[0,1]$ interval of real numbers. In the discussion that follows, this choice makes the interpretation of results more straightforward. The fuzzy sets initially symmetrically segment each variable's domain but progressively depart from the initial shape by training (figure 5).

The method used yielded satisfactory accuracy levels. The overall accuracy was 80% and per-class accuracies varied from 67% (forest) to 100% (water). Classification accuracies for dense and sparse urban, were 75 and 81%, respectively. Because the method yields some unclassified instances, it is not possible to construct the complete accuracy matrix but only its diagonal. An equally important finding

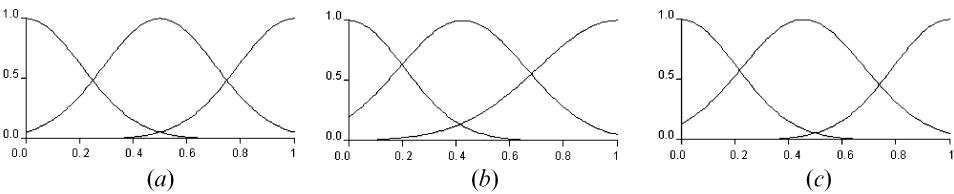


Figure 5. The x -axis displays the normalized variable value. The y -axis shows membership function. Small, medium and large fuzzy sets correspond to the three bell-shaped curves from left to right, respectively. (a) Fuzzy sets before training. (b) Trained fuzzy sets for DEM and (c) slope.

that will be discussed here is the interpretability of the method. The GNN classifier is characterized by a concrete relationship between the rule set and the neural network skeleton and vice versa. This property enables us to actually draw the neural network skeleton. Visualization is a key aid towards the understanding of how input variables contribute to classification of patterns.

The system is trained to infer the 10 most efficient rules. The fuzzy sets are forced to always overlap and their relative order is kept. Rule weights are not used. The learning rate used is 0.01 and the maximum number of epochs is set to a number (here 2000) rarely reached since a solution is found quite early in the training process. The choice of those parameters is experimental and subject to ongoing research. The system is then pruned and only seven (for six land use classes) rules are retained. To demonstrate the process of inferring the network skeleton by the rule base, we can start by building the network for the forest rule (figure 6) in the set which is:

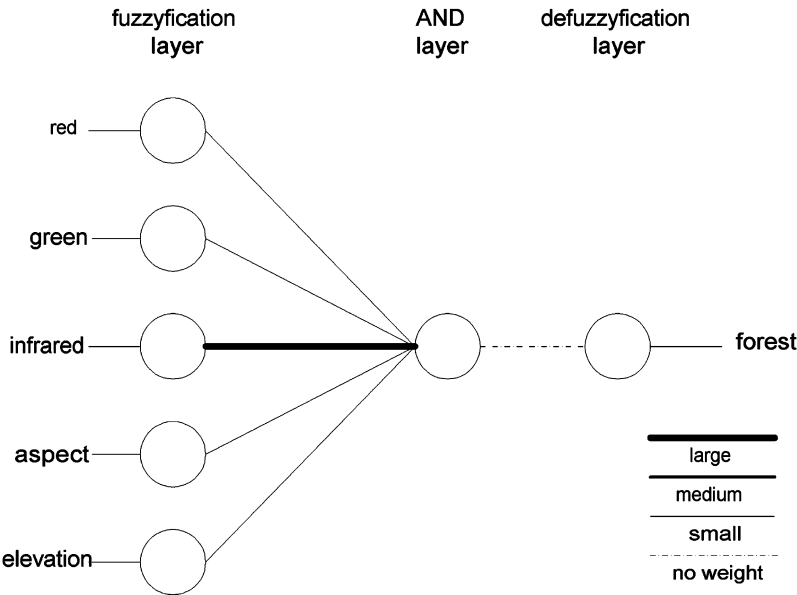


Figure 6. Conversion of rule 1 into a neural network skeleton. The thickness of links connecting fuzzyfication to the AND node denote the exact fuzzy set used. The rest of the links are not assigned to fuzzy sets.

RULE 1:

```

if
  IRS Red Band is small
  and IRS Green Band is small
  and IRS NIR Band is large
  and Aspect is small
  and Elevation is small
then
  Forest

```

The defuzzified values can be converted to a crisp output vector, using for example the winner takes all rule. However, if we avoid this last step, we should be able to use the values as a means of non-linear spectral un-mixing to be applied on the mixed pixels (Foody 2000, Baraldi *et al.* 2001, Qiu and Jensen 2004). Specifically, an original output vector, obtained by the GNN and assigned to the crisp class agricultural, is something like:

(urb_dense|0, urb_sparse|0.1, forest|0.1, agricultural|0.7, soil/sparse_veg|0.1, water|0)→agricultural

However, since the numbers associated with each category denote the coexistence of categories other than the dominant one, we could decide to keep that information by using it for pixel unmixing in future studies.

As a convention we use three progressively thicker line widths to show the fuzzy set that is actually assigned to the specific link. In this model we use the same linguistic terms for all input variables. It is always small, medium and large. We could use the same metaphor, the progressive line thickness, even with different fuzzy sets. The semantics would then be slightly different, representing the intensity, in other words the rank, of the variable and not the exact tag itself. Nevertheless, the visual value of this coding holds. The symbolization is not applicable for fuzzy sets that are not ordered. The full rule set is given as a fuzzy table (Zadeh 1997) in table 1 and the converted rules to the NN presented in figure 7.

An additional comment on figure 7 is that urban sparse defuzzification node is backwards-connected to two AND nodes. The weights have values of 1 and the maximum value of the two is transmitted forward. The node is labeled 'max' in the diagram to denote this fact. The NN view is a very clear picture of what is actually happening during the classification process.

4. Results

A first comment on the classification is that non-spectral information proves to be equally valuable in the separation of land use types. If spectral information was

Table 1. Final, after pruning, fuzzy table. Contains seven rules, two rules for urban sparse (top) and one rule for each of the other categories.

IRS red	IRS green	IRS IR	Aspect	Elevation	Land use
Large	Small	Small	Small	Small	Urban dense
Small	Large	Small	Small	Small	Urban sparse
Small	Small	Small	Large	Small	Soil/sparse vegetation
Small	Small	Large	Small	Small	Forest
Small	Small	Large	Small	Large	Agricultural
Medium	Medium	Small	Small	Small	Urban sparse
Small	Small	Small	Small	Small	Water

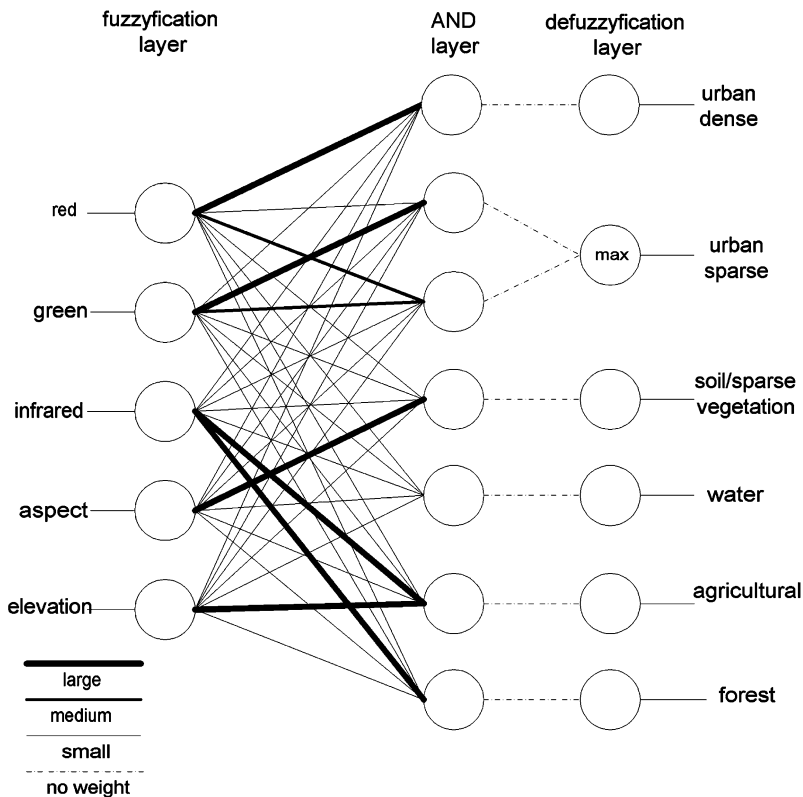


Figure 7. Neural network view. The number of nodes in the fuzzyfication (input) layer equals the number of input variables in the input vector after pruning. The number of nodes in the defuzzification (output) layer equals the number of the target land cover classes. The number of nodes in the AND (hidden) layer equals the number of rules in the rule set (table 1). Weights assigned to links connecting fuzzyfication to AND nodes are actually fuzzy sets. Link thickness corresponds to a specific set as shown by the legend. Weights assigned to links connecting AND to defuzzification nodes are always 1 to maintain interpretability. Maximum activation is selected in case two rules lead to the same output class as for urban sparse.

sufficient, the pruning procedure would lead to the removal of non-spectral nodes in the input layer. In contrast, we see that aspect, for example, is a dominant factor in classifying as soil sparse vegetation cover. This conforms to previous knowledge of the area which suggests that a particular aspect range is indeed subject to erosion because it is exposed to very strong local winds. This finding is encouraging the use of contextual information in general to increase dimensionality and include otherwise difficult to classify categories in the schema. Contextual information other than topographic could be demographic, meteorological, geological or other depending on the case at hand.

Second, as expected, infrared reflectance is clearly used to separate the vegetation classes, forest and agricultural, from the rest. What is more interesting is that the differentiation between the two, forest and agricultural, is done on the basis of elevation. It seems that vegetation at high elevations is characterized as agricultural whereas high infrared reflectance at low and medium elevations is classified as forest. This might imply that certain types of agricultural cultivations that are more

common in lower altitudes, such as olive trees, are not present in the training sample. This is very useful information in order to include more such samples and further refine the classification in the future.

Third, we see that one rule is not enough to separate the urban sparse category. There are two AND nodes, hence two rules, leading to this category. This conforms well to our perception that in the specific area urban sparse is actually composed of two subcategories. The first subcategory contains urban green areas. The second subcategory refers to urban areas which exhibit lower density, compared with the urban dense category, and at the same time have more vegetation than the core city. The green band is deployed to separate the first subcategory. The second subcategory is characterized by medium green reflectance and red reflectance that is an order of magnitude less than that of urban dense. This is an example of one rule being not enough to adequately partition the decision space (figure 8) and the way GNNs handle heterogeneous classes.

Fourth, one rule is enough for all other classes except urban. This means that a very simple structured NN with just one hidden layer and not fully connected can perform the classification. The performance might further improve by trying different configurations with respect to fuzzy set shape, fuzzy sets number per class, test and training size and mixture as well as other parameters (Dixon 2004).

5. Conclusion

In conclusion, the implementation of GNN provide a means of combining diverse sources of information, here spectral and non-spectral, to increase dimensionality

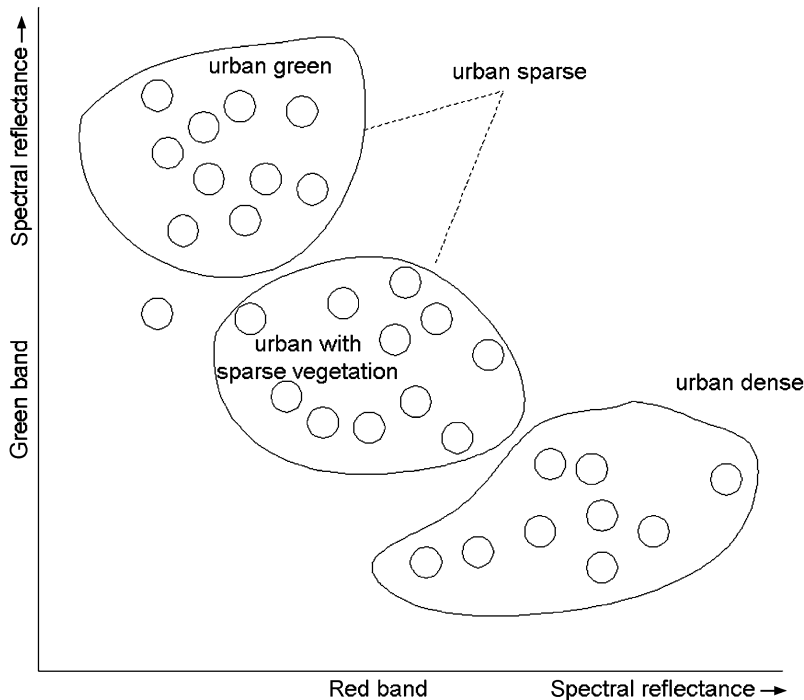


Figure 8. Urban sparse class contains two distinct sub-classes. This results in two separate rules in the rule base leading to the same output. Converting to network skeleton we realize two AND nodes leading to the same output node.

and better segment decision space. In the future, other kinds of contextual information could be used to perform difficult classification tasks that cannot be resolved purely on the basis of spectral information. The fact that we are living in an era where spatial digital content is not a sparse commodity any more makes this possibility even more appealing.

Moreover, the high level of interpretability of the granular neural networks allows the investigation of causality connecting input variables to output classes. The transparency of granular neural networks can be used to interpret the semantics of the deducted rules in the rule set. It is possible to create an accurate representation of the rule set as a neural network with one hidden layer. This representation provides us with an insight into the actual process taking place which in turn increases our confidence in the model and makes refinement easier.

In addition, the use of a metaphor to symbolize weights assigned on each link connecting input variables to the hidden layer (AND layer) is evaluated. This symbolization further promotes transparency providing an easy comprehensible view of the neural network model. The visual inspection of the exact fuzzy sets and the contribution to each particular fuzzy rule provides a very useful synopsis of how classification operates. This can prove to be a significant aid in understanding complex problems.

References

- ATKINSON, P.M. and TATNALL, A.R.L., 1997, Introduction to neural networks in remote sensing. *International Journal of Remote Sensing*, **18**, pp. 699–709.
- BARALDI, A., BINAGHI, E., BLONDA, P., BRIVIO, P. and RAMPINI, A., 2001, Comparison of the Multilayer perceptron with neuro-fuzzy techniques in the estimation of cover class mixture in remotely sensed data. *IEEE Transactions on Geosciences and Remote Sensing*, **39**, pp. 994–1005.
- BENEDIKTSSON, J.A. and BENEDIKTSSON, H., 1999, Neuro-fuzzy and soft computing in classification of remote sensing data. In *EUROPTO Conference on Image and Signal Processing for Remote Sensing V*, 1999, Florence, Italy, SPIE, 3871, pp. 176–187.
- BENEDIKTSSON, J.A. and SVEINSSON, J.R., 1997, Feature extraction for multisource data classification with artificial neural networks. *International Journal of Remote Sensing*, **18**, pp. 727–740.
- BORTOLAN, G., 1998, An architecture of fuzzy neural networks for linguistic processing. *Fuzzy Sets and Systems*, **100**, pp. 197–215.
- BORTOLAN, G. and PEDRYCZ, W., 2002, Linguistic neurocomputing: the design of neural networks in the framework of fuzzy sets. *Fuzzy Sets and Systems*, **128**, pp. 389–412.
- BURROUGH, P.A. and McDONELL, R.A., 1998, *Principles of Geographical Information Systems*, p. 190 (New York: Oxford University Press).
- CONGALTON, R.G. and GREEN, K., 1998, *Assessing the Accuracy of Remotely Sensed Data: Principles and Practices*, p. 160 (Boca Raton, FL: CRC Press).
- DIXON, B., 2004, Applicability of neuro-fuzzy techniques in predicting ground-water vulnerability: a GIS-based sensitivity analysis. *Journal of Hydrology*, **4**(2), pp. 1–38.
- ERSOY, O.K. and HONG, D., 1990, Parallel self-organizing, hierarchical neural networks. *IEEE Transactions on Neural Networks*, **1**, pp. 167–178.
- FOODY, G., 2000, Estimation of sub-pixel land cover composition in the presence of untrained classes. *Computers & Geosciences*, **26**, pp. 469–478.
- FOODY, G., 2002, Status of land cover classification accuracy assessment. *Remote Sensing of Environment*, **80**, pp. 185–201.
- HAN, J., LEE, S., CHI, K. and RYU, K., 2002, Comparison of neuro-fuzzy, neural network, and maximum likelihood classifiers for land cover classification using IKONOS

- multispectral data. In *IGARSS, International Geoscience & Remote Sensing Symposium*, 2002, Toronto, Canada, Sponsored by the IEEE, Vol. 6, pp. 3471–3473.
- ITO, Y. and OMATU, S., 1997, Category classification using a self-organizing neural network. *International Journal of Remote Sensing*, **18**, pp. 829–845.
- KANELLOPOULOS, I. and WILKINSON, G.G., 1997, Strategies and best practice for neural network image classification. *International Journal of Remote Sensing*, **18**, pp. 711–725.
- LIN, C.-T., LEE, Y.-C. and PU, H.-C., 2000, Satellite sensor image classification using cascaded architecture of neural fuzzy network. *IEEE Transactions on Geosciences and Remote Sensing*, **38**, pp. 1033–1043.
- LIPPMANN, R.P., 1987, An introduction to computing with neural nets. *IEEE ASSP Magazine*, **2**, pp. 4–22.
- MITRA, S., PAL, S. and MITRA, P., 2002, Data mining in soft computing framework: a survey. *IEEE Trans. Neural Networks*, **13**, pp. 3–14.
- NAUCK, D., 2003, Fuzzy data analysis with NEFCLASS. *International Journal of Approximate Reasoning*, **32**, pp. 103–130.
- NAUCK, D. and KRUSE, R., 1999, Neuro-fuzzy systems for function approximation. *Fuzzy Sets and Systems*, **101**, pp. 261–271.
- NAUCK, U., 1999, Design and implementation of a neuro-fuzzy data analysis tool in Java, Thesis, Technische Universität Braunschweig, Institut für Betriebssysteme und Rechnerverbund, Aufgabenstellung und Betreuung Prof. Dr. R. Kruse Braunschweig. Partly available online without registration at: <ftp://fuzzy.cs.uni-magdeburg.de/pub/nefclass/manual.pdf> (accessed 1 August 2005).
- PEDRYCZ, W. and REFORMAT, M., 2005, Genetically optimized logic models. *Fuzzy Sets and Systems*, **150**, pp. 351–371.
- PEDRYCZ, W. and VASILAKOS, A., 1999, Linguistic models and the linguistic modeling. *IEEE Transactions on Systems, Man and Cybernetics*, **29**, pp. 745–757.
- PEDRYCZ, W. and VUKOVICH, G., 2001, Granular neural networks. *Neurocomputing*, **36**, pp. 205–224.
- QIU, F. and JENSEN, J.R., 2004, Opening the black box of neural networks for remote sensing image classification. *International Journal of Remote Sensing*, **25**, pp. 1749–1768.
- RUMELHART, D.E., HINTON, G.E. and WILLIAMS, R.J., 1987, Learning internal representations by error propagation. In *Parallel Distributed Processing-Explorations in the Microstructure of Cognition*, RUMELHART, D.E., MCCLELLAND, J.L. and the PDP Research Group (Eds), Vol. 1, pp. 318–362 (The MIT Press: Boston, MA).
- SHAPIRO, A.F., 2002, The merging of neural networks, fuzzy logic, and genetic algorithms. *Insurance: Mathematics and Economics*, **31**, pp. 115–131.
- Shuttle Radar Topography Mission (SRTM) 2000, Available online at: <http://www2.jpl.nasa.gov/srtm/> (accessed 1 August 2005).
- WANG, L.X. and MENDEL, J.M., 1992, Generating fuzzy rules by learning from examples. *IEEE Transactions on Systems, Man and Cybernetics*, **22**, pp. 1414–1427.
- ZADEH, L.A., 1965, Fuzzy sets. *Information and Control*, **6**, pp. 338–353.
- ZADEH, L.A., 1975, The concept of a linguistic variable and its application to approximate reasoning-I. *Information Sciences*, **8**, pp. 199–249.
- ZADEH, L.A., 1978, Fuzzy sets as a basis for a theory of possibility. *Fuzzy Sets and Systems*, **1**, pp. 3–28.
- ZADEH, L.A., 1997, Toward a theory of fuzzy information granulation and its centrality in human reasoning and fuzzy logic. *Fuzzy Sets and Systems*, **90**, pp. 111–127.
- ZADEH, L.A., 1999, From computing with numbers to computing with words—from manipulation of measurements to manipulation of perceptions. *IEEE Transactions on Circuits and Systems I*, **45**, pp. 105–119.