KNOWLEDGE BASED EXPERT SYSTEMS IN REMOTE SENSING TASKS: QUANTIFYING GAIN FROM INTELLIGENT INFERENCE

M. Shoshany

GeoInformation Engineering, Faculty of Civil & Environmental Engineering, Technion, Israel Institute of Technology, Haifa 32000, Israel - maximsh@tx.technion.ac.il

Commission VII, WG VII/6

KEY WORDS: Knowledge Base, Reasoning, Theory, Fusion, Classification, Data Mining

ABSTRACT:

How to measure gain from the use of intelligent inference ? How can the complexity of the recognition/ classification task be estimated? What is the type of evidence which best suits an .inference mechanism? These questions are addressed here in their theoretical and methodological aspects. Their practical implications are demonstrated with 'real' crop mapping task. For this purpose, simple rule-based system is compared with expert system based on Dempster – Shafer evidential reasoning algorithm. The advantage of using 'soft' / implicit evidence with the Dempster-Shafer algorithm over the use of 'hard' / explicit evidence with decision – tree type procedure is discussed.

1. INTRODUCTION

Spatial, temporal and spectral complexity of remote sensing recognition tasks necessitates the use of Knowledge Based Expert Systems (KBES). Such complexity concern the fact that the same surface phenomenon may emerge in different ways in imagery sources (see discussion for example in Lu and Weng,(2007).). These systems facilitate algorithmic adjustments of the use of classification or recognition rules according to the local context. KBES combine available evidence, procedural knowledge regarding priorities in implementing evidences in parallel or sequentially; and an inference mechanism. The procedural knowledge may be constructed based on human expert knowledge or through extensive learning of the recognition problem at hand. Data Mining and Knowledge Discovery are technologies suitable for conducting such extensive learning. In parallel, there are numerous inference mechanisms available which shift the emphasis into designing efficient expert systems considering different strategies for information extraction and processing. Such strategies also concern the efforts made in the search for evidence on the one hand and the information gain from using this evidence on the other hand. Recently, there is growing interest in the field of information seeking and utilization with the adoption of Evolutionary – Ecological Models of Foraging (Pirolli and Card,1999). Comparison between strategies based on assessment of information gain versus information cost is a central element in the Information Foraging Theory. A primary problem in adopting this theory in remote sensing concern quantitative evaluation of gain from evidence versus gain from inference (Shoshany and Cohen,2007). The presentation in the conference would be divided into three elements: review of fundamental terms, concepts and strategies in the implementation of KBES, assessment of the similarity between information foraging and remote sensing tasks; and finally, discussion of the ways to estimate gain from evidence and inference in mapping missions.

2. FUNDAMENTAL TERMINOLOGY

First, it is important to recognize that at the initial stage of the recognition task we are dealing with propositions, or better say multiple propositions for each recognition task. Evidence is then defined (Wikipedia) as “any objectively observable or demonstrable circumstance which tends to indicate or disprove a proposition". Two main types of evidence can be discerned:

Explicit Evidence: refers to the "notion of plain distinct expression that leaves no need to infer" (Merriam-Webster). In the remote-sensing narrow band signatures of specific materials and very unique combinations of spectral reflectance thresholds are examples of explicit evidence. The production of such evidence would typically require exhaustive search (see methods proposed by Peddle and Ferguson,2002).

Implicit Evidence: type of proposition represents a conclusion "capable of being understood from something else though unexpressed: capable of being inferred" (Merriam-Webster).

Implicit evidence would consist of generalizations, associations and contextual information representing some level of non-uniqueness and conflicts with reference to the object subject to recognition (e.g., Cohen and Shoshany,2005). While the production of such evidence might be relatively easy, its use would require inference capable of overcoming such non-uniqueness and conflicts.

Considering that remote sensing evidence would be usually located on a continuous scale inbetween the extreme explicit to extreme implicitness, it is clear that the role of inference is elementary. Inference is defined (Merriam-Webster) as "the act of passing from one or more propositions, statements, or judgments considered as true to another the truth of which is believed to follow from that of the former". Deductive, inductive, abductive, analogical or common-sense reasoning would facilitate such inference (Durkin,1994). The selection of type of evidence and inference is a crucial step in developing KBES strategies.

3. KNOWLEDGE BASED EXPERT SYSTEMS STRATEGIES

Two fundamental extreme types of inference mechanisms are utilised in Artificial Intelligence :

- Domain Dependent Inference (DDI): is suitable for relatively explicit body of evidence in which the conclusion is entailed and no, or low level of inference is required. DDI inference mainly controls the combination order of the different evidences and is represented in the procedural knowledge. The sophistication in expert systems based on DDI is embedded within them during their rules' formation rather than in their functioning way. The construction of procedural knowledge and evidential basis requires heavy information analysis, learning procedures and feature assessment. Rule based expert systems implementing binary decision trees are extreme examples of DDI and are widely used in remote sensing (e.g., Goodenough et al.,1994; Kartikeyan et. Al.,1995; Chan et al.,1999). GeoAIDA (Tonjes et al.,1999) is a good example of such strategy, where expert system based on semantic networks was developed implementing specific (explicit) evidence, including contextual information. The inference mechanism applied was then a procedural sequential decision tree type. Other extremely different example of this strategy concerns the use of Classification Tree Analysis (CTA) techniques and their recent improvements with Boosting and Bagging methods (e.g., Lawrence et. al.,2004). Again in this method there is maximal exploitation of the information content in the data and the (computerized) construction of a decision tree specifically (explicitly) applicable to the data at hand.

- Domain Independent Inference (DII): is suitable for relatively implicit body of evidence. The DII inference relate to their associated characteristics: the relative belief, support, certainty and weight, rather the information sources themselves and thus is independent of any specific recognition or decision making problem. The combination of evidence would then be based on general deductive, inductive or abductive procedures (Durkin,1994). A significant element of the intelligence represented by DII is embedded within it during the development of the generalized inference algorithm which is independent of any specific recognition or decision making problem, and usually adopted without investing in such development. The sophistication of the system which is embedded in the inference methodology is then dependent on the inference capabilities to overcome non-uniqueness and conflicts (e.g., Cohen and Shoshany,2005).

In reality there is rather a mixture of strategies tailored according to information sources and expert systems tools available. However, there is relatively little comparison made between DDI and DII based strategies in terms of their performance and the types of evidences used. Performance assessment of these strategies in areas which differ from the training areas is of special interest. Would it be better in terms of 'cost' and performance to deepen the search for domain explicit evidence or to broaden the implicit evidential basis? These same questions are addressed in "Information Foraging".

4. THE INFORMATION FORAGING PERSPECTIVE

Charnov,(1976) and Stephens and Kerbs,(1995) proposed the foraging theory to explain foraging behavior and strategies in nature. This theory is based on the observation that animals evaluate the availability of food sources and at the same time the efforts required in order to consume them. Prey searching efforts, competition with other predators and energy required to catch the prey are among the main ' costs' of obtaining the food. The amount of food (prey size) gained and its quality represents the gain from these efforts. 'Within Patch' searching strategies concern exploiting the food available at a certain area before migrating to another patch. While such strategy minimizes the search and migration energy, there is decrease in food availability and maybe increase competition. In the remote sensing context such strategy would be represented by limiting the information search for one data source: visible, IR or microwave spectral bands. Domain Dependent Inference (DDI) strategies are also a type of 'Within Patch' foraging behavior. 'Between Patch' strategies represent large area search for patches of high food availability or low competition. Its implementation requires gathering of larger amounts of data and better skills in optimizing their use. Animals adopting these strategies will have higher resilience to changing conditions and would reduce the potential degradation from overgrazing. In the remote sensing context 'Between Patch' strategies would be characterized by implementing multiple sources (e.g., multi-spectral; multi-temporal; multi-resolution). The uses of Domain Independent Inference (DII) strategies are analogous to this later animal behavior.

'Translating' the foraging theory into the information and knowledge extraction arena required quantitative treatment of the information gained from the efforts made for searching it. Such treatment is crucial for assessing the relative performance obtained from the implementation of the different strategies. Pirolli and Card,(1999) provided the following simplified expression for assessing relationships between the time invested in producing the information items and the information gain.

$$G = g\, T_b / t_b \qquad (1)$$

Where, T_b is the total amount of processing time, t_b is the average processing time per information item and g is the average information gain per item.

In the next section the treatment of information gain from developing and implementing expert systems in remote sensing is presented.

5. ESTIMATING RECOGNITION ENERGY AND GAIN FROM INFERENCE

The complexity of the recognition task is determined by the level of confusion between the different surface phenomena as their appear in the multi-temporal and /or multi-spectral and or multi-resolution feature space. In other words: highly separable classes would not require much work in recognizing them and vice versa. There are numerous unsupervised classification algorithms which may facilitate 'automatic' determination of classes in the feature space. The inherent separability which exists inbetween these classes is inversely related to the "effort" needed to be invested in constructing the knowledge base.

Transformed Divergence (TD) is one of the most used measures of separability which therefore may facilitate the estimation of total effort required by the KBS to reach high level recognition. It is determined for each pair of classes according to the following formula (Swain and Davis,1978):

$$TDij = 2(1 - (\exp(- Dij/8))) \qquad (2)$$

where

$$Dij = 0.5\, tr\left((C_i - C_j)(C_i^{-1} - C_j^{-1})\right) + 0.5 tr\left((C_i^{-1} - C_j^{-1})(\mu_i - \mu_j)(\mu_i - \mu_j)^T\right)$$

where:

i and j = the two signatures being compared
Ci = the covariance matrix of signature i
$_i$ = mean vector of signature i
tr = the trace function (matrix algebra)
T = transposition function

The technical advantage of using TD stems from the fact that it provides an expected threshold value for high separability (=2000). The recognition energy (Re) required by the expert system to resolve the existing level of inseparability (unresolved complexity) can be estimated by:

$$Re = \sum (2000 - TD_i) \qquad (3$$

Where i is the index for pairs of classes.

Similarly, the effort needed or actually invested in producing an expert system for resolving the existing level of inseparability at a certain level of expected / obtained accuracy (*Ue*) may be estimated by:

$$c \cdot \sum (K_W + I_W + G_W) = Re \cdot Ue \qquad (4)$$

Where *Kw* and I_W are the work/efforts invested in producing the KB and formalizing the procedural inference mechanism, respectively (i.e., working hours, manpower units etc.); *c* is a scaling/calibration coefficient translating effort units to productive recognition energy ($Re \cdot Ue$); and, G_W is the productive recognition / information gain from using the inference without investing effort in developing it: either by using DII or by using previously developed DDI. Different combinations of K_W, I_W and G_W facilitating the same recognition targets thus represent the relative sophistication embedded in each of the inference mechanisms.

While it is possible to estimate the work/ effort invested in extracting evidence (*Kw*) or in constructing procedural mechanisms of inference (I_W), it is impossible to determine directly the 'effort' made by deduction/ induction/ abduction. Equation 9 suggests that it might be estimated in productive recognition energy ($Re \cdot Ue$) units when it is treated as an element of the overall work invested in building the recognition system.

Shoshany and Cohen,(2007) implemented this approach by assessing the efforts in constructing the evidential basis and the recognition results obtained for a simple Rule Based system (DDI) and for Dempster- Shafer based KBS (DII). This assessment was conducted within the framework of a crop mapping task in Mediterranean region where in evidences were derived for one study area and then the two expert systems were implemented for another study area which was not trained. Figure 1 shows that Rule based systems may have an advantage when workloads area high and when there is no differentiation in efforts required for constructing thematic type of evidence (implicit evidence) and continuous type of evidence (explicit evidence). As the differences in evidence production between explicit and implicit increases the advantage of the DII is enhanced.

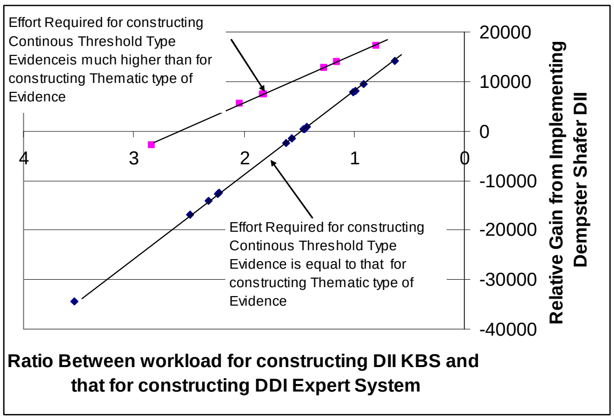

Figure 1: Gain from Dempster Shafer Type Inference as obtained for different work loads (efforts required for constructing evidence).

6. DISCUSSION AND CONCLUSIONS

This study presented one of the first attempts ever made in comparing the gain from evidence versus that from inference, which are two central elements in the reasoning the remote sensing recognition process. Assessment of relationships between them is extended when considering implicit versus explicit evidence and domain-dependent versus independent inference. Domain-independent inference (DII) represents one of the common intelligence capabilities founded on general principles of induction, deduction or abduction. The highest level of expert system 'intelligence' is gained in resolving most complex problems from a most redundant data/information of the implicit type. The bottom line suggests that DII may have significant gain when the production of implicit evidence requires a third or less effort than that required for producing hard explicit evidence. However, it must be emphasized that the gain identified here for DII in general terms must be attributed specifically to the Dempster-Shafer theory of evidence. In our opinion it facilitates an important shift from a search for the principle attributes/components/ordinates to approaches integrating vast amounts of implicit data by adjusting the context and by being resilient to contradicting and heterogeneous evidence.

REFERENCES

Chan, J. A., Chan, K.., and Yeh, A.G. ,1999, Detecting the nature of change in an urban environment: a comparison of machine learning algorithms, *Photogrammetric Engineering & Remote Sensing,* 67, 213-225.

Charnov, E. L.,(1976). Optimal foraging: The marginal value theorem. Theoretical Population Biology, 9, 129-136.

Cohen, Y., and Shoshany, M.,2002, Integration of remote sensing, GIS and expert knowledge in national knowledge-based crop recognition in Mediterranean environment. *Int. J. of Applied Earth Observation and GeoInformation*, 4, 75-87.

Cohen, Y., and Shoshany, M.,2005. Analysis of convergent evidence in an evidential reasoning knowledge-based classification. Remote Sensing of Environment, 96, 518-528.

Durkin, J.,1994, Expert Systems: Design and Development, Prentice Hall, Englewood Cliffs, NJ, 800 p.

Goodenough, D.V., Charlebois, D., Matwin, S., and Robson, M.,1994, Automatic Reuse of Software for Expert System Analysis of Remote Sensing Data. *IEEE Transactions on Geoscience and Remote Sensing, 32(1), 525-533.*

Kartikeyan, B., Majumder, K.L., and Dasgupta, A.R.,1995, An Expert System for Land Cover Classification. *IEEE Transactions on Geoscience and Remote Sensing, 33(1), 58-66.*

Lawrence, R., Bunn, A., Powell, S., Zambon, M.,2004, Classification of remotely sensed imagery using stochastic gradient boosting as a refinement of classification tree analysis. *Remote sensing of Environment*, 90, 331-336.

Lu, D., and Weng, Q.,2007, A Survey of image classification methods and techniques for improving classification performance. Int. J., of Remote Sensing, 28(5), 823-870.

Peddle, D.R., and Ferguson, D.T.,2002, Optimisation of multisource data analysis : an example using evidential reasoning for GIS data classification. *Computers & Geosciences*, 28, 45-52.

Pirolli, P., and Card, S. K.,1999, Information Foraging, *Psychological Review*, 106(4), 643-675.

Shoshany, M., and Cohen, Y.,2007, Evidence versus Inference in Remote Sensing Recognition. *International Journal of Remote Sensing* 28(11), 2613-2634.

Stephens, D. W. & Krebs, J. R.,(1986). Foraging theory . Princeton, NJ: Princeton University Press.

Tonjes, R., Growe, S., Buckner, J., Liedtke, C-E.,1999, Knowledge-Based Interpretation of Remote Sensing Images Using Semantic Nets. *Photogrammetric Engineering and Remote Sensing*, 65(7), pp. 811-821.

Smith, J.,1987a. Close range photogrammetry for analyzing distressed trees. *Photogrammetria*, 42(1), pp. 47-56.

Names of journals can be abbreviated according to the "International List of Periodical Title Word Abbreviations". In case of doubt, write names in full.