In journal: Fundamenta Informaticae, 30 (3/4), pages 227-240, (1997)
Also available as Tech Report TR-93-00, Dept CS, Univ. Ottawa, Canada, (1993)
http://www.cs.utexas.edu/users/pclark/papers

Improving Image Classification by Combining Statistical, Case-Based and Model-Based Prediction Methods

Peter Clark, Cao Feng, Stan Matwin

Ottawa Machine Learning Group
Computer Science
University of Ottawa
Ottawa K1N 6N5, Canada
{pclark,cfeng,stan}@csi.uottawa.ca

Ko Fung

Canada Centre for Remote Sensing
588 Booth St., Ottawa K1A 0Y7
Canada
fung@ccrs.emr.ca

Abstract. Evidence for image classification can be considered to come from two sources: traditional statistical information derived algorithmically from image data, and model-based evidence arising from previous expertise and experience in a given application domain. This paper presents a study of classification techniques based on both these sources (traditional algorithmic and model-based), and illustrates how they can be combined. A prototype image classification system, called Cabaress, has been constructed which implements these methods. We evaluate Cabaress as applied to the problem of identifying crops in agricultural fields, based on classifying image segments extracted from radar image data. Our results demonstrate this mixed-method approach can achieve improved classificational accuracy.

1 Introduction

Following the successful launch of RADARSAT, the research activities in the application of synthetic aperture radar (SAR) have been increased significantly. Synthetic apreture radar is a microwave device which, as a remotely sesensing device, has some advantages over the optical sensors. The marked contribution of the SAR over optical sensor is its capability to penetrate cloud and its sensitivity to moisture contents. However the speckle characteristics, the inherent single frequency illumination and the sensitivity of the SAR device to topography makes SAR a difficult device to use for plant species identification. Despite much research devoted to automatic SAR image classsification, results obtained from the classification of SAR images are inferior when compared to multiple channels optical image classification results.

At the Canada Centre for Remote Sensing, an investigation team led by Dr. Ronald Brown has been investigating the application of radar images to agriculture. It has been found that using the traditional classification methods, the classification acccuracy from the radar images is usually lower than that of the multiple channel optical images with similarly spatial resolution. A rule based approach was attempted in addition to the parametric classification approach, using the crop rotation rules generally practised by most farmers at the test sites. The result was found to be promising [1].

Following this direction, our goal in this research is to further improve image classification by using *mixed-method classification techniques*. In particular, we have drawn on classification technologies from statistics, expert systems and case-based reasoning, and have integrated them in a system called Cabaress[1], the subject of this paper. As each classification technique differs in the source of evidence it uses, their combination offers the potential of further improving accuracies achieved by individual methods in isolation. In this paper, we describe how these methods are integrated and demonstrate how their combination achieve small but statistically significant accuracy improvements compared with each method individually.

2 Overview

The problem we are addressing is that of *image description classification*, namely given a high-level description of an image, identify which of a set of mutually disjoint classes the image is in. Note that we do not classify the image directly, but instead base classification on a set of features computed from it.

In the crop classification task considered here, the images to be classified are *image segments*, i.e. regions of an image whose pixels have similar characteristics. Each image segment covers the area of a different agricultural field, extracted from a large SAR image using field boundaries from a Geographic Information System (GIS). The image segment description consists of values of a set of attributes (e.g. mean horizontally-polarised backscatter, averaged over all image pixels; perimeter of the area to classify; geographical co-ordinates of the centroid), and the classification is one of seven possible crops which the field contains (peas, wheat, barley etc.).

In this application, evidence for classifying new image segment descriptions can be considered to come from two sources:

1. **Spectral:** Given a set of directly observable features in the new image segment, and a set of training examples (segments with known classifications), we can infer the most likely class of the new segment. We draw on two techniques to do this:

   (a) **Bayesian Discrimination:** Using Bayesian techniques, we can estimate the correlation between each observed feature and the possible classes from a set of training data, and then use this information to label the most likely class of the new segment.

   (b) **Case-Based Reasoning:** By identifying which known image segment(s) the new candidate segment is most similar to, we can use their classes to assign a class to the candidate.

2. **Model-Based:** In addition to evidence from the image segment itself, there may be expertise available about the application domain to assist in classification. Such expertise is based on a *model* of relationships in the application domain, and hence we refer to this as model-based evidence.

   The expertise underlying the domain model can be derived using two methods; either through face-to-face interviews with human experts, or automatically

[1]Case-BAsed REmote Sensing Shell

identified using machine learning techniques (e.g. [2, 3, 4]). In the machine learning approach, which we have used in this research, learning programs are applied to large bodies of data about the domain in order to identify patterns and regularities. These regularities, encoded in a computationally manipulable way, can then be used by a classification system to assist in classification decisions.

In our crop identification task, the domain model comprises of knowledge about crop-rotation patterns and is automatically constructed by a learning system applied to a database of crop rotation histories. For example, the domain model expresses the fact that a field is rarely left as fallow for two years in a row, and that canola will probably be planted in the year after a field has been left fallow. Given the histories of crops planted in each field, such model-based evidence can contribute to the classification.

Our objective is to integrate these classification technologies to improve image classification. First we describe in detail each of these three different knowledge sources individually, and how they can be applied to classify image segment descriptions. We then overview Cabaress, a generic system which integrates these three methods together. Finally, we provide a detailed evaluation of the system as applied to the problem of crop identification from synthetic aperture radar (SAR) images.

3 Individual Sources of Classification Knowledge

3.1 The Case-Based Knowledge Source

In the field of Artificial Intelligence (AI), case-based reasoning (CBR) has now become a key form of problem-solving (eg. [5, 6]). While the paradigm is not new, it is only recently that CBR technology has become widely studied and its advantages better understood.

Case-based reasoning asserts that much problem-solving involves identifying and reasoning from specific, similar problems which have already been encountered and solved. This assertion constrasts with the assumption underlying expert systems, namely that problems can be solved by purely abstract reasoning, removed from specific previous experiences. Aha et al. describe a simple, general framework summarising the key steps in case-based reasoning [7]:

1. **Similarity Assessment:** Identify cases in the case-base (i.e. a a set of previously encountered problems with known solutions) which are most similar to the new case (i.e. the current problem) to be solved.
2. **Case Retrieval:** Retrieval of those cases from the case-base.
3. **Solution Transfer/Adaptation:** Apply (maybe with some adaptation) the old solution(s) for the retrieved case(s) to solve the new case.

In the context of image classification, the problem is to classify a new image and the solution is simply the assignment of a class label to that image. Case-based classification thus requires identifying images most similar to the one to be classified, and then using them to identify and assign the most likely class to it. In addition to achieving good classificational accuracies, CBR provides explanations to the user based on previous cases, preferred by many users.

Cabaress uses a simple instantiation of the case-based paradigm, using a k-most-similar-neighbour algorithm to identify the most similar cases, i.e. using Euclidian distance in case-space as the similarity metric. k is specified by the user[2], and refers to the number of cases to retrieve from the case-base for classifying the new image description. We refer to this selection method as 'k-most-similar-neighbour'.

If only a single case is retrieved (ie. k=1), then the classification of that case is assumed to be the most likely class of the new case. If more than one is retrieved, then a probability vector is constructed based on the frequency of each class's occurrence in the retrieved sample (size k). The probability vector comprises a list of each possible class and its associated probability of occurence. For example, if $k = 10$, and the ten retrieved cases are in classes `[bar,fal,bar,flx,wht,bar,cgs,bar,wht,wht]`, then the probability vector will look `[bar-0.4,wht-0.3,fal-0.1,flx-0.1,cgs-0.1]`. If the case-based knowledge source is used alone, the most probable class in this vector is then assigned to the new case.

The case-based source normalises the data, so the values for each attribute are scaled and shifted to have a mean of zero and a standard deviation of one. For discrete attribute values, the distance betweeen two cases with different values of a discrete attribute is taken to be 1.0 (ie. equivalent to one standard deviation distance of numeric attributes), and 0.0 if they are the same. Unknown attribute values are replaced with the mean (numeric) or most common (discrete) value of that attribute.

3.2 The Bayesian Knowledge Source

In addition to case-based methods, standard methods from statistics have been developed for image classification. To compare/combine this style of image classification with other methods, we use a standard *maximum likelihood classification* (MLC) algorithm based on application of Bayes' rule. This approach has been used successfully in image classification (e.g. [1, 8]) and other classification tasks (eg. [9, 10, 11]).

For each prediction class C_i, the probability that an example with a feature vector $V = v_1, ..., v_n$ belongs to C_i is:

$$p(C_i|V) = \frac{p(V|C_i) \times p(C_i)}{p(V)} \tag{1}$$

$$= \frac{p(V|C_i) \times p(C_i)}{\sum_j (p(V|C_j) \times p(C_j))} \tag{2}$$

[2]In fact, the user specifies the percentage of training cases to use and k is then calculated from this.

where:

$p(C_i|V)$ = probability that the example is of class C_i, given it has features V
$p(V|C_i)$ = probability that an example of class C_i will have feature vector V
$p(C_i)$ = prior probability for class C_i.

The values of $p(V|C_i)$ are assessed using a set of training examples, with known classification. Ideally, $p(V|C_i)$ is directly based on observations of the relative frequencies of co-occurence between V and C_i. However, if the number of dimensions of the vector V increases and not enough training examples are collected, $p(V|C_i)$ can be estimated either by making an independence assumption (when the features are relatively uncorrelated) or by using an n-dimensional Gaussian probability density function, described by a mean vector and covariance matrix estimated from the training data [12].

3.3 The Model-Based Knowledge Source

In addition to information directly available from the image, there may be extra knowledge about the domain which can be applied to improve classificational accuracy. In the case of the SAR crop prediction task, there are many possible sources of expertise which could play a role:

- **Crop rotation patterns:** Most farmers practise crop rotation to increase their yields. The rotational patterns, combined with information about crops planted in the previous years, can be used to predict the current crop.
- **Climate information:** The geographical location and the nearby topography of a field provides information about the likely climate and hence likely crops to be planted in the region (eg. crops preferring warmer climate will be rarer in the north).
- **Geographical information:** Economical constraints affect likelihoods of crops. For example, crops with a high transportation cost and low profit margin may become less probable the farther away from a storage silo the field is; knowledge about the type and contours of the terrain may improve the accuracy in crop identification. Data from Geographic Information Systems (GIS) could potentially provide a rich source of additional knowledge to assist in crop classification.
- **Financial information:** Farmers also base their decisions about which crops to plant based on market potentials, aiming to maximise profitability. Information about expected crop prices and likely future demand could again assist in classification.
- **Crop 'portfolio management':** To reduce exposure to financial risk, farmers maintain a 'portfolio' of crops in their fields. Knowledge about which fields are owned by which farms, and how such portfolios are managed, could again contribute to improving classificational accuracy.

- **Agricultural information:** Knowledge about local soil types and conditions could be used to help predict likely crops to be planted.

- **Advice centres:** Knowledge of recommendations from influential financial or agricultural advice centers will affect farmers' decisions about which crops to plant.

To allow encoding of this sort of information, Cabaress contains a rule-based inference engine which we refer to as the model-based knowledge source. The inference engine takes a knowledge-base, encoded as a set of '*if...then...*' production rules, and applies it to new image segment descriptions to generate a probability vector expressing the likelihood of different classes for the new segments. This 'expert system' architecture has been used extensively in artificial intelligence applications, and provides a flexible means for encoding domain knowledge [13, 14].

In our prototype knowledge base, we have only encoded knowledge about the first item above, namely crop rotation patterns, but of course the reasoning infrastructure could be used with knowledge about the other aspects mentioned above. Two examples of crop rotation rules in the system are:

```
IF   last year's crop = barley
THEN current crop =
        barley (31%),
        canola (25%),
        peas (16%),
        fallow (14%),
        wheat (8%),
        flax (4%),
        canary grass (2%).

IF   last year's crop = barley
AND  the crop two years ago = canola
THEN current crop =
        wheat (44%),
        barley (29%),
        peas (14%),
        flax (7%),
        fallow (3%).
```

The first of these is called a **one-year** rule, as it bases its prediction on the one-year history of the field. The second rule is similarly called a two-year rule.

A model of rotational patterns, to be encoded in rules such as these, can be extracted in two ways: either using machine learning/pattern recognition algorithms applied to a database of crop histories; or from manual interviews with experts. In our application we used the former method, applying a learning algorithm which simply collects statistics on different crop rotational patterns which are present in the database of crop histories. The output of this learning system is encoded as rules such as those just presented.

As well as encoding these rotational patterns as rules, two standard ways of presenting such patterns are using **transition matrices** and **probabilistic state transition graphs** [8]. A transition matrix can be used to show, for a given crop in year $Y-1$, what the probability of the crop in year Y will be (e.g. see Table 1). A probabilistic state transition graph can be used to depict the same data in graphical format, with the thickness of the arcs corresponding to the state transition probability. We use these representations to illustrate parts of the rotational model extracted from the data sets we are using.

3.3.1 One Year Rotational Rules

The transition matrix for one-year rotation rules (ie. given the crop in year $Y-1$, predict the crop in year Y) which was extracted is shown in Table 1. The same data is presented in the probabilistic state transition graph shown in Figure 1, which nicely illustrates the likely transitions. These transitions were converted into rotational rules of the form described earlier, to be used by the inference engine during classification of new examples.

3.3.2 N-Year Rotational Rules

In a similar way, transition rules based on the previous N years' crops can be extracted from the data. From the two year transition matrix, a transition graph can be sketched which is sensitive not only to the previous year's crop but also to the crop before that. An example is shown in Figure 2. Here, only the most probable ($p \geq\approx 25\%$) transitions are depicted. The most probable transition from a node (ie. a crop) depends not only on that crop but also on the crop before that. To illustrate this, several points are associated with a node, each representing different ways in which that node could have been arrived at. Thus, for example, `barley` is most likely to follow the sequence `canary` →`peas`, but `wheat` is most likely to follow `wheat` →`peas`.

Prev. crops Y-1	**→**	**No. Exs**	**wht**	**pea**	**can**	**cgs**	**bar**	**flx**	**fal**
wht	→	(129)	25	13	13	3	25	3	13
pea	→	(69)	34		7	10	39	5	2
can	→	(89)	51	13	1		29	2	2
cgs	→	(17)	17	47	17	11			5
bar	→	(135)	8	17	26	2	27	5	11
flx	→	(25)	32	12	4		44		8
fal	→	(39)	25		69		5		

Table 1: Transition Matrix for One Year Rotational Rules (probabilities in %).

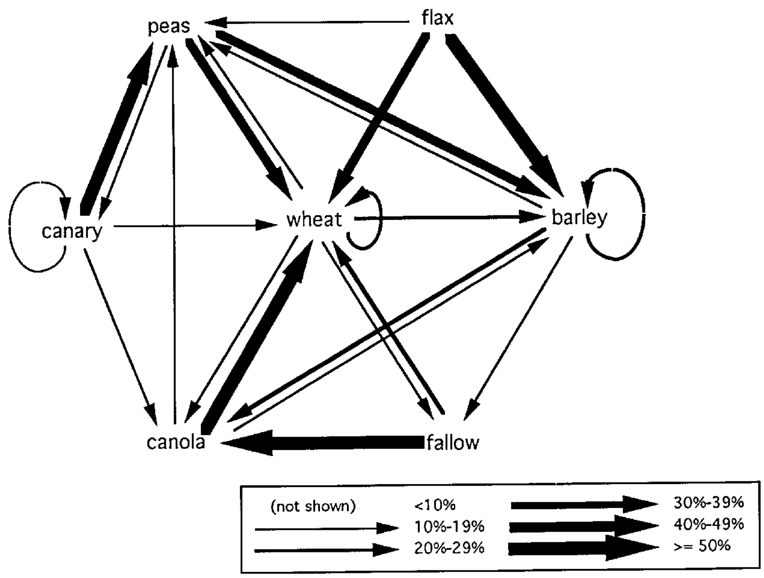

Figure 1: Probabilistic state transition graph for one-year rules. A node represents the state of a crop field in some year, defined solely by the crop currently planted (eg. peas). An arc points to a likely subsequent state of that field in the next year, where the thickness of the arc denotes the likelihood. State transitions are non-deterministic, ie. there are several possible transitions from a given state, as the current crop alone is not sufficient to determine the subsequent state precisely.

Paths represent probable rotational patterns, eg. `peas` →`canary` →`peas` →`barley`. The main findings from the transitions extracted from the data, shown in this graph, are as follows:

- The graph is still non-deterministic, for example a pattern `barley` →`peas` could be followed by either `wheat`, `canary` or **barley** with roughly equal probability.
- Transitions from `wheat` are almost completely unpredictable (even given knowledge of crops in previous years) hence only a single point is associated with `wheat`.
- The network is in fact two disjoint connected graphs:
  - the main network
  - A single loop, repeating the sequence `barley` →`barley` indefinitely. Note that this loop is separate because the graph only considers histories up to two years old. Accounting for a longer history would cause this loop to merge with the rest of the network.

Rules such as these can act as a rotational model in the domain. We now examine how

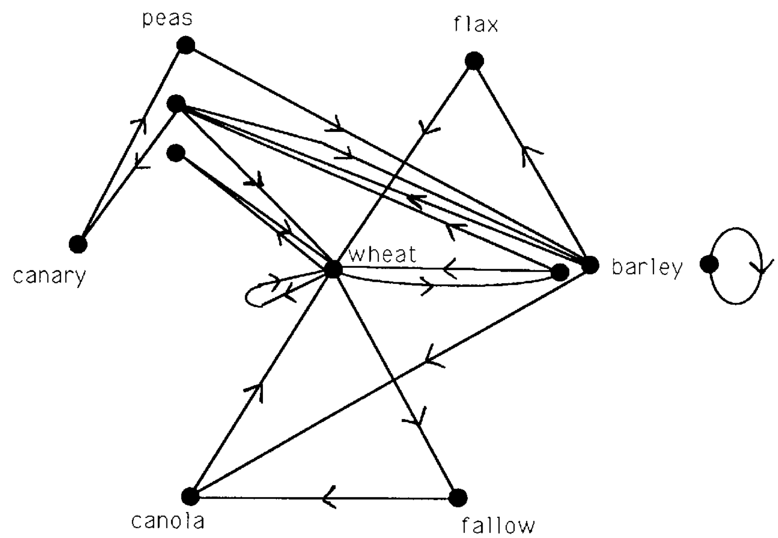

Figure 2: Probabilistic state transition graph for two-year rules. A node represents the state of a crop field in some year, defined by the crop currently planted (shown in the node's label) *and* the crop planted in the previous year. An arc points to a likely ($\geq$ 0.25 probability) subsequent state of that field in the next year. Again, state transitions are non-deterministic as the two-year history is still not sufficient to completely predict the subsequent state.

model-based predictions derived from these rules can be combined with the prediction methods using spectral evidence described earlier.

4 Combining the Knowledge Sources

Each of the three knowledge sources described takes as input an image description, and produces as output a probability vector expressing the probabilities of different classes being the correct classification of the image. These vectors are combined by Cabaress using a weighted average:

$$p(C_i) = w_{cbr} p_{cbr}(C_i) + w_{bayes} p_{bayes}(C_i) + w_{rules} p_{rules}(C_i)$$

where p_{cbr}, p_{bayes} and p_{rules} are the probabilities from the case-based, Bayesian and model-based knowledge sources respectively and $w_{cbr} + w_{bayes} + w_{rules} = 1$. This is depicted in Figure 3, showing the overall architecture of the complete system.

The relative magnitude of the coefficients w_{cbr}, w_{bayes}, and w_{rules} express the relative confidence/reliability to be placed in each of the three knowledge sources. These values are normally selected automatically, using a standard cross-validation technique.

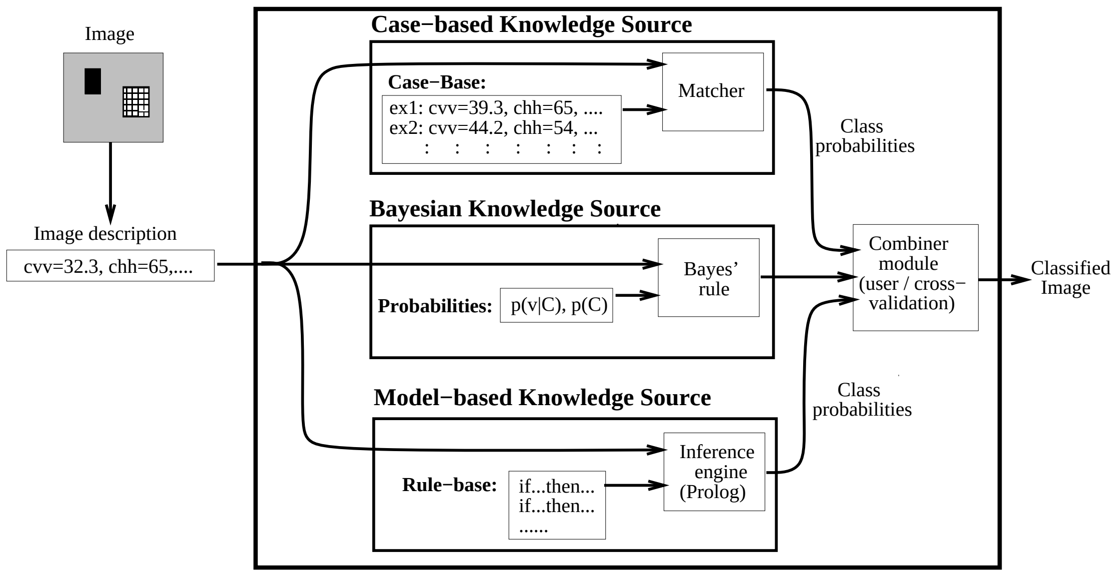

Figure 3: Overall architecture of Cabaress.

In this method:

1. The space of possible weight assignments is systematically explored by sampling different points within it. Each point corresponds to a particular choice of weights. An estimate of Cabaress's future performance with this choice of weights is found as follows:
   (a) The training data is partitioned into ten equally sized groups $g_1, ..., g_{10}$.
   (b) For each group g_i:
      i. Train the system using all groups *except* g_i (ie. use these nine groups as the case-base, and compute the Bayes' probabilities from these nine groups).
      ii. Test the system using g_i as a testing set.
   (c) Take the average of the ten results thus obtained.
2. Select the choice of weights from step 1 which produced the highest estimate of future performance. This corresponds to finding the maximum of the performance surface in the weight space.

For example, Cabaress might sample the following points in weight space (using intervals of 0.5), and obtain the following performance estimates from this cross-validation procedure:

Point: $(w_{cbr},\ w_{bayes},\ w_{rules})$	Performance estimate
(0, 0, 1)	49%
(0, 0.5, 0.5)	53%
(0, 1, 0)	48%
(0.5, 0, 0.5)	56%
(0.5, 0.5, 0)	54%
(1, 0, 0)	53%

and hence select weights $(w_{cbr},\ w_{bayes},\ w_{rules}) = (0.5,\ 0,\ 0.5)$ as the best to use, as they produced the highest expected performance from the cross-validation procedure.

In practice, Cabaress samples the space at a finer grain (in our experiments, an interval of 0.1 was used). Note there are only two, not three, degrees of freedom as $w_{cbr} + w_{bayes} + w_{rules} = 1$. In addition if one knowledge source is removed from consideration (as in our experiments where we wish to directly compare two knowledge sources, Section 5), then only one degree of freedom remains.

From standard statistical theory, cross-validation offers a probabilistic guarantee that the resulting weighted combination of sources will be at least as good as a single source: in other words, with a high probability (based on the degree of sampling performed during cross-validation), performance during cross-validation will approximate performance during future trials. This is based on the assumption that the original training data used in cross-validation, and future test data, are randomly drawn from the same population, and hence overall performance on each should also be the same when averaged over many such trials – though of course performance on particular training/testing samples may vary. In the case where two of the knowledge sources contain erroneous or random information, cross-validation will assign them weights at or near zero, thus placing all the classification weight on the third reliable knowledge source.

In addition, as an alternative, Cabaress also allows the user to select the weights manually, based on his/her knowledge of the data and rules and experimental trials.

5 Experimental Evaluation

5.1 The Classification Task

Our objective in this paper is to integrate classification technologies from artificial intelligence and statistics, and assess their individual and combined performances. In this section, we present the results of a detailed evaluation of Cabaress, the system integrating these different classification methods.

We evaluated Cabaress on the problem of crop prediction from SAR data, in which the images to be classified are image segments, extracted from a larger SAR image using boundary information from a geographical information system. The performance task is as follows: given SAR image segments of an agricultural field, and a description of the crops planted in that field in previous years, predict the crop currently in the field. Such predictions are important for assessing the likely crop productions in the area studied (Section 1).

We evaluated Cabaress using two data sets, one for agricultural fields in 1989 and one for 1990. Each data set contains 301 examples, each example being a description of a single field. The description contained six attribute-value pairs:

1. The mean pixel intensity of the image segment recording the horizontally polarised SAR backscatter from the field (`chh_m`).
2. The mean pixel intensity of the image segment recording the vertically polarised SAR backscatter from the field (`cvv_m`).
3. The crop currently planted in the field (`c`).
4. The crop planted in the field in the previous year (`pc`).
5. The crop planted in the field two years previously. (`p2c`).
6. The crop planted in the field three years previously (`p3c`).

Additional information about the images and the field was also available (e.g. geographical location, standard deviations of the mean pixel intensities). However, an earlier study indicated that this extra information did not contribute significantly to classificational accuracy and so was not used in this experiment [16]. There are seven types of crops, resulting in seven mutually exclusive classes to predict. These are: wheat, barley, flax, fallow, canola, canary grass, and peas.

5.2 Experimental Method

Experiments were conducted on each data set (1989, 1990) in turn. The data was split randomly, 67% used to train the system and 33% used to test the system.

The three different knowledge sources were trained using the training data as follows:

Case-Based: The training data simply forms the case-base. These cases are matched with each test example during testing to find the most similar cases.

Bayesian: Values of $p(V|C_i)$ are computed from the training examples using the appropriate method as described in Section 3.2.

Model-Based: The model of crop rotations is generated automatically from the known crop histories (i.e. not using the image segment statistics of horizontally and vertically polarised backscatter). To do this, all the data (training and testing) was used, *except* information about the current crop of the testing examples was kept hidden when generating the model. (A new model was generated for each separate experimental run).

Note that it would be misleading to generate the rotation model using all the data available, as this will feed information about the (supposedly unseen) test data into the system (e.g. Cabaress may note the rotational pattern `a` $\rightarrow$ `b` $\rightarrow$ `c` $\rightarrow$ `d` from a single test example, then when classifying this same test example given the crop history `a` $\rightarrow$ `b` $\rightarrow$ `c` $\rightarrow$ `?` it will necessarily predict correctly). To avoid this, we hide data in the 'current crop' column of the test data set during model generation. For interest, we also compared results with domain models generated when this data was not withheld, using the data sets in their entirities (results shown in the last columns of Tables 2 and 3).

We compared performances of the knowledge sources using different quantities of historical information, by controlling a 'history window' on the datasets:

0yr Classify purely on the two image-specific attributes (`chh_m` and `cvv_m`). Note the model-based knowledge source cannot be applied if no crop history is available.

1yr For case-based/Bayesian classification alone, classify using three attributes (the two image attributes and the previous crop (`chh_m`, `cvv_m` and `pc`).
For model-based classification alone, generate and apply a model of one-year rotations. The one-year rotation model is generated using all the data (except the current crop of test examples). The rules are applied by looking at the single attribute `pc` in the test examples.
For case-based/Bayesian and model together, we train and apply the case-based/Bayesian classifier using just the two image attributes `chh_m` and `cvv_m` (ie. not `pc`). The rotational model is applied as before, ie. using the attribute `pc` in the test examples. A cross-validation method was used to automatically determine the best weighted combination of evidence from the case-based/Bayesian and model-based knowledge sources.

2yr For case-based/Bayesian classification alone, classify using the four attributes `chh_m`, `cvv_m`, `pc` and `p2c`.
For model-based classification alone, generate and apply using a model of two- and one-year rotations. Two-year rotational rules are tried first on the test examples. If none apply, then one-year rules are tried (and if still none apply, a default most-likely crop is assigned to the test example).
As before, for case-based/Bayesian and model together the case-based/Bayesian classifier uses just `chh_m` and `cvv_m`, and the model is applied using the `pc` and `p2c` of the test examples.

3yr Same as two year, except also using the three year history attribute `p3c`.

The experiments were repeated 50 times each, using different random splits of the data. The percentage accuracies were recorded, and the standard errors (denoted by $\pm$) computed from the samples.

5.3 Results and Discussion

The results are shown in Tables 2 and 3. These show the results of experiments on the 1989 data set combining case-based and model-based knowledge sources (top of Table 2), combining Bayesian and model-based knowledge sources (bottom of Table 2), and the same for the 1990 data set (Table 3).

Columns one and two show the accuracies achieved by each knowledge source alone, and column three their combination. For interest, column four shows the combination again, but using all the data (including the crop of the test examples) to generate the crop rotation model, as discussed earlier.

Classification of 1989 SAR data

Case-based & model-based knowledge sources, separately & combined:

History window	case-based	model-based	case- and model-based combined	(With 'ideal' model, gen. from all data)
0yr	52.0 ±1.0			
1yr	50.6 ±0.8	34.4 ±0.4	53.7 ±0.4	(62.1 ±0.1)
2yr	45.5 ±1.0	32.0 ±0.4	53.5 ±0.4	(62.3 ±0.6)
3yr	42.2 ±0.9	32.9 ±0.4	54.2 ±0.4	(68.6 ±0.5)

Bayesian & model-based knowledge sources, separately & combined:

History window	Bayesian	model-based	Bayes' and model-based combined	(With 'ideal' model, gen. from all data)
0yr	49.0 ±0.5			
1yr	48.8 ±1.0	34.4 ±0.4	50.3 ±0.5	(52.1 ±0.6)
2yr	48.9 ±0.9	32.0 ±0.4	49.6 ±0.4	(53.1 ±0.4)
3yr	48.6 ±1.1	32.9 ±0.4	50.7 ±0.5	(60.0 ±0.5)

Table 2: Accuracies (%) on 1989 SAR data. (The 1989 images are lower quality than the 1990 images, hence the overall accuracies are lower).

Classification of 1990 SAR data

Case-based & model-based knowledge sources, separately & combined:

History window	case-based	model-based	case- and model-based combined	(With 'ideal' model, gen. from all data)
0yr	74.1 ±1.0			
1yr	72.9 ±0.9	38.2 ±0.4	75.2 ±0.4	(77.3 ±0.2)
2yr	69.7 ±0.8	46.1 ±0.4	75.4 ±0.4	(78.1 ±0.3)
3yr	68.2 ±1.1	43.4 ±0.4	75.7 ±0.4	(82.2 ±0.4)

Bayesian & model-based knowledge sources, separately & combined:

History window	Bayesian	model-based	Bayes' and model-based combined	(With 'ideal' model, gen. from all data)
0yr	74.0 ±0.7			
1yr	74.7 ±0.8	38.2 ±0.4	75.7 ±0.4	(76.3 ±0.3)
2yr	74.3 ±1.1	46.1 ±0.4	75.7 ±0.4	(77.2 ±0.4)
3yr	74.1 ±0.9	43.4 ±0.4	75.4 ±0.4	(81.9 ±0.3)

Table 3: Accuracies (%) on 1990 SAR data.

These experiments produced several interesting and striking results. The most important result is that the combined classifier system produces small (around 1%-2%) but statistically significant improvements in accuracy, despite the fact we have encoded just one of the many possible sources of domain knowledge in the model-based prediction module (Section 3.3). This illustrates that knowledge about the problem domain, external to that computable from image segments themselves, can usefully contribute to classification. We hope that encoding additional information about the domain can further improve classificational accuracy.

A second, and surprising, finding was that the accuracy of the Bayesian and case-based knowledge sources alone actually degraded as more history information was made available to them. This is surprising as one would intuitively expect more information to produce improved accuracy. The most likely explanation is that both these classifiers treat attributes in a simple manner, assuming attribute independence – however, for crop rotation patterns it is precisely the order dependence of attributes which provides the predictive power, and which these classifiers on their own will thus fail to capture. The result illustrates that simply obtaining extra information for classification is not sufficient – it's use must also be modelled appropriately, as we have sought to do with the model-based knowledge source.

A third result is that, while one-year old historical information improves classificational accuracy, there is no evidence that additional historical information (two- and three-year history) contributes further to classificational accuracy of the combined case-based/Bayesian and model-based system. Again, this is surprising as extra information normally would be expected to improve accuracy. Two factors may contribute to the contrary result. First, the size of the training data is relatively small for generating reliable two-year rotational rules. Given 7 classes, there are 7^2 possible two-year rules. Given approximately four years of crop histories of 301 examples, this results in around 600 three-year sequences, i.e. resulting in just $600/7^2 \approx 12$ examples per rotational rule with which to estimate class probabilities. The improvement in accuracies using the 'ideal' model (the last column in Tables 2 and 3), which approximates perfect knowledge about likely rotation patterns, shows that more robust classification may be possible if better rotational rules were available. Second, there is still considerable uncertainty in crop rotation patterns: even given two (and three) year histories, there is still considerable variation in the crop the farmer will choose to plant next. It thus appears that there is still considerable scope for expanding the model-based component to take into account the additional factors which the farmers consider.

6 Conclusions

We have presented a framework for integrating model-based and statistical image classification techniques, and described and evaluated a system, Cabaress, which implements such an integration. Our evaluation on SAR crop prediction, using a limited domain model of crop rotation patterns, shows that combining these types of classification knowledge can improve classificational accuracy. In addition, the different knowledge sources offer improved explanation facilities for justifying classificational decisions, the system being able to refer to both specific, previous images which look similar to the current image (case-based reasoning) and provide causal explanations of why one particular classification is more likely than another (model-based prediction).

In our particular application domain, the experiments also revealed some interesting properties of the crop prediction task. In particular, while crop rotation patterns improve accuracy, it is clear there is still additional information besides crop histories used by farmers when planning which crops to plant (e.g. geographical constraints, financial considerations). An interesting future direction would be to expand the model-based knowledge source to account for these additional factors, to further improve the system's classification decisions and provide enhanced explanation facilities. The architecture of Cabaress, presented in this paper, offers a generic shell in which such extensions can now be easily developed.

Acknowledgements

The authors would like to thank Dr. R. Brown and Mr. D. Bedard for supporting this development with their advice on the crop rotation rules and the SAR images and GIS files. This paper is funded by the Canada Centre for Remote Sensing and is supported by the University of Ottawa.

References

[1] D. Bedard, R. Brown, J. Naunheimer, B. Brisco, and S. Huang. Investigating crop rotation knowledge to improve agricultural sar classifications. In *15th Canadian Symposium on Remote Sensing*, pages 239–244, 1992.

[2] J. Ross Quinlan. Learning efficient classification procedures and their application to chess endgames. In J. G. Carbonell, R. S. Michalski, and T. M. Mitchell, editors, *Machine Learning, vol. 1.* Tioga, Palo Alto, CA, 1983.

[3] Peter Clark and Tim Niblett. The CN2 induction algorithm. *Machine Learning Journal*, 3(4):261–283, 1989.

[4] R. Michalski, I. Mozetic, J. Hong, and N. Lavrac. The multi-purpose incremental learning system AQ15 and its testing application to three medical domains. In *AAAI-86*, volume 2, pages 1041–1045, CA, 1986. Kaufmann.

[5] Agnar Aamodt and Manuela Veloso, editors. *Proc. International Conference on Case-Based Reasoning (ICCBR'95).* AcknoSoft, 1995.

[6] M. Keane, editor. *Proceedings Second European Workshop on Case-Based Reasoning (EWCBR'94).* AcknoSoft, 1994.

[7] David W. Aha, Dennis Kibler, and Marc K. Albert. Instance-based learning algorithms. *Machine Learning*, 6(1):37–66, 1991.

[8] Hans Middelkoop and Lucas L. F. Janssen. Implementation of temporal relationships in knowledge-based classification of satellite images. *Photogrammetric Engineering and Remote Sensing*, 57(7):937–945, 1991.

[9] Peter Clark and Tim Niblett. Induction in noisy domains. In I. Bratko and N. Lavrač, editors, *Progress in Machine Learning (proceedings of the 2nd European Working Session on Learning)*, pages 11–30. Sigma, Wilmslow, UK, 1987.

[10] D. Michie. Bayes, turing and the logic of corroboration. In *Machine intelligence and related topics: an information scientist's weekend book*. Gordon & Breach, NY, 1982.

[11] C. Feng, A. Sutherland, R. King, S. Muggleton, and R. Henery. A comparison of classification algorithms in machine learning, statistics and neural networks. In *Proc. 1992 Conference on AI and Statistics*, 1992.

[12] P. M. Mather. *Computer Processing of Remotely-Sensed Images*. Wiley, Chichester, 1987.

[13] J. L. Alty and M. J. Coombs. *Expert Systems: Concepts and Examples*. NCC, Manchester, UK, 1984.

[14] B. Andrews. *Successful Expert Systems*. Financial Times Business Info Ltd., London, 1989.

[15] Peter Clark. Representing knowledge as arguments: Applying expert system technology to judgemental problem-solving. In T. R. Addis and R. M. Muir, editors, *Research and Development in Expert Systems VII*, pages 147–159. Cambridge Univ. Press, 1990. (Also available as TIRM-90-043 from Turing Institute, 36 N Hanover St, Glasgow, UK).

[16] Peter Clark and Stan Matwin. A study of case-based and model-based techniques for sar crop prediction, and their integration into a general case-based expert system. Technical report, Dept. of Computer Science, 1992. (Report for the Canadian Centre for Remote Sensing).