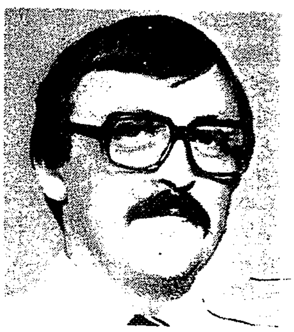
Rule induction in forensic science

Ian W Evett
Statistician
Central Research Establishment
Home Office Forensic Science Service
U.K.

E J Spiehler
Consultant
Newport Beach, California
U.S.A.

Glass fragments are frequently found when forensic scientists examine the clothing of a person who is suspected of a crime such as house-breaking, and it is possible to determine the elemental composition and refractive index of very small fragments. It is useful if the scientist can use these data to determine whether an unknown fragment is of a particular type of origin. We have used the rule induction package BEAGLE to determine whether reasonable classification procedures can be derived from a given database of analyses of glass samples of known origin. The application of BEAGLE is described and its performance is compared with that of two conventional statistical techniques. It is clear that rule induction can provide fresh perspectives on classification problems of this nature.



The principal author worked for several years as an operational forensic scientist before moving to the Central Research Establishment to take responsibility for information/computer facilities within the Forensic Science Service (FSS). He has spent three periods outside the FSS, including a year as a main grade statistician at the Welsh Office. He presently holds a grade 7 research post concerned with developing effective methods for evaluating forensic science evidence. His work includes planning artificial intelligence research within the field.

Presented at KBS IN GOVERNMENT : Online Publications, Pinner, UK, 1987

Introduction

Glass is a material which figures prominently in the investigation of crimes such as burglary and criminal damage in which it is common for a window to be smashed violently, either to gain access or as an act of vandalism. If a suspect is apprehended for such an offence then it is almost a routine matter to submit articles of his clothing to a forensic science laboratory so that a scientist may determine whether or not there is evidential material present. If the scientist finds fragments of broken glass on the clothing then it is possible to measure certain of their properties, even if they are little larger than sugar grains. One of the questions which the scientist addresses relates to the class of origin of the fragments, for example: are they window glass? or are they more likely to have come from a smashed bottle or beer glass? To assist in answering such questions, the forensic science laboratories maintain data collections on glass samples of known types. Determining the means of using such a collection to greatest effect is not a trivial problem.

It is the role of the Central Research Establishment of the Home Office Forensic Science Service to enhance the effectiveness of the UK operational forensic science laboratories through a wide range of support and research services. As part of that role we have a commitment to ensure that advances in information technology are, where appropriate, exploited for the benefit of the forensic scientist. We have several lines of research and development both in-house and in collaboration with external organisations. One of those lines has been to evaluate selected low cost artifical intelligence software packages including expert system shells and rule induction programs.

This paper describes the application of one of those packages, called BEAGLE, to the classification of glass fragments. The forensic problem will first be explained in a little more detail and the conventional statistical approaches will be outlined. BEAGLE will then be described using glass classification for illustration. The performance of BEAGLE will be compared with that of the conventional methods and, finally, some of the advantages and disadvantages of rule induction will be discussed.

Glass Classification

The most useful property of glass for forensic purposes is its refractive index (RI). Not only does RI vary greatly from one glass object to another but also it is possible to carry out precise measurements on very small fragments. When, therefore, glass fragments are found as a result of examining clothing it is a matter of routine to determine the RI of at least a sample of them. Another stage of the examination, generally requiring somewhat larger fragments, is multielement analysis. This is done by means of a scanning electron microscope; much more expensive than RI determination, it will normally be carried out on only a few fragments in a given case. Interpretation of the results of such analyses clearly requires a background information collection of some sort and the easiest way of building such a collection is to accumulate casework data. In each case there will normally be at least one control sample of glass from a known source and, apart from its function in that case, the data from the analyses can be stored for future use. Data on over 200 glass samples have been collected in this way at the Home Office Forensic Science Laboratory, Birmingham. The important matter to resolve is how to use the data most effectively to determine the type of origin of an unknown glass fragment.

One has only to reflect briefly to realise that there are many different types of glass objects in use and catering for them all could make the job of classification rather complicated. For the purpose of this study we settled for a simple scheme. The most important distinction is between window glass (which includes vehicle windows) and non-window glass. Most modern window glass is made by a process which involves floating the molten glass on a pool of molten tin and this, not surprisingly, is known as 'float glass'; our classes for window glass were, then, 'float' and 'non-float'. The non-windows were split into 'containers', 'tableware' and 'headlamps'. An extract from the collection is shown in Table 1.

There are well established statistical methods for undertaking classification in this sort of situation. The easiest to understand is the 'nearest neighbour' method which has a simple basis. The position of a point on a geographical map can be defined by two numbers which describe its latitude and longitude. The distance between two points can be measured with a rule and converted according to the appropriate scale. The nine variables (RI plus eight elements) determined for a glass sample can, by analogy, be regarded as defining a point on a nine dimensional map and it is a straightforward task to calculate the 'distance' between two samples by repeated application of Pythagoras' Theorem. Given the data for an unknown sample we can search the collection to find the closest members: the unknown is classified according to the class of its nearest neighbours. If the closest samples are all float window glass then the unknown is classified as float window; if they are all from containers then the unknown is classified as container glass. When the nearest neighbours are of two or more types then the classification can be assigned according to the majority vote.

TABLE 1

AN EXTRACT FROM THE DATABASE

RI	Na	Mg	Al	Si	K	Ca	Ba	Fe	TYPE
1.51761	13.98	3.60	1.36	72.73	0.48	7.83	0.00	0.00	WF
1.51784	12.68	3.67	1.16	72.11	0.61	8.70	0.00	0.00	WF
1.51966	14.77	3.75	0.29	72.02	0.03	9.00	0.00	0.00	WF
1.52667	13.99	3.70	0.71	71.57	0.02	9.82	0.00	0.10	WF
1.51808	13.43	2.87	1.19	72.84	0.55	9.03	0.00	0.00	WF
1.51811	12.96	2.96	1.43	72.92	0.60	8.79	0.14	0.00	WNF
1.51730	12.35	2.72	1.63	72.87	0.70	9.23	0.00	0.00	WNF
1.51820	12.62	2.76	0.83	73.81	0.35	9.42	0.00	0.20	WNF
1.52725	13.80	3.15	0.66	70.57	0.08	11.64	0.00	0.00	WNF
1.52151	11.03	1.71	1.56	73.44	0.58	11.62	0.00	0.00	C
1.51969	12.64	0.00	1.65	73.75	0.38	11.53	0.00	0.00	C
1.51666	12.86	0.00	1.83	73.88	0.97	10.17	0.00	0.00	C
1.52369	13.44	0.00	1.58	72.22	0.32	12.24	0.00	0.00	C
1.52058	12.85	1.61	2.17	72.18	0.76	9.70	0.24	0.51	C
1.51969	14.56	0.00	0.56	73.48	0.00	11.22	0.00	0.00	T
1.51115	17.38	0.00	0.34	75.41	0.00	6.65	0.00	0.00	T
1.51602	14.85	0.00	2.38	73.28	0.00	8.76	0.64	0.09	H
1.51623	14.14	0.00	2.88	72.61	0.08	9.18	1.06	0.00	H
1.52065	14.36	0.00	2.02	73.42	0.00	8.44	1.64	0.00	H

The compositions are expressed in weight percent of the respective oxides.

Sample Types:
- WF - Float Window
- WNF - Non-float Window
- C - Container
- T - Tableware
- H - Headlamp

This is not always clear cut, particularly when the nearest neighbours are some distance away, but there is no need to go into such details here. For this study we used the three nearest neighbours for classification.

Another widely used statistical technique is called discriminant analysis. This employs certain assumptions about the way in which the various properties vary across each of the classes. If those assumptions are valid then, in theory, discriminant analysis will give the best performance. However, in the problem at hand it is probable that the assumptions do not hold and so the performance may not be very good. Discriminant analysis establishes a number of fairly complicated mathematical functions - one for each of the classes in the collection. When an unknown sample is tested its data are substituted into each of the functions and the classification essentially follows from the one which gives the largest value. The calculations were carried out for us by a colleague using the Statistical Package for the Social Sciences (SPSS).

BEAGLE

The title is an acronym: Bionic Evolutionary Algorithm Generating Logical Expressions. Its creator is Richard Forsyth who markets the package through a company called Warm Boot Ltd, Nottingham. It is radically different in concept from conventional statistical methods, belonging to the branch of artificial intelligence which is known as machine learning. Some approaches have their roots in psychology and attempt to model the way in which human beings acquire knowledge, but BEAGLE works in a way which mimics in several ways the natural process of evolution.

To operate BEAGLE it is necessary to give it a database containing examples of known types and a simple goal which, in our application, could be something of the kind 'find rules for deciding whether a sample is window glass or not'. Inspection of Table 1 shows that magnesium levels tend to be higher in windows than non-windows which suggests the simple rule 'if the magnesium level is greater than zero it is a window'. This, however, will lead to errors because the table shows that some non-windows have a mesureable magnesium content. So it might be worth refining the rule by including, for example, aluminium: 'if the magnesium is greater then zero and the aluminium is less then 1.5% it is a window'. This rule could then be checked and it might be found that it correctly classified a few more of the non-windows but at the cost of misclassifying a few more windows. Again the rule could be modified, or another rule could be created to be used in conjunction with the first, and the performance re-assessed. For the human this rapidly becomes tedious and confusing but it is a resonable task for a properly programmed computer.

BEAGLE follows this sort of procedure with subtle refinements of its own. Once we give BEAGLE the data and the target then we can, if we wish, suggest some starting rules. If we don't make any suggestions then the program will make some up by taking random combinations of the variables and arithmetic and logical operators: in fact, it starts with a list of ten such rules and checks their performance on the database. Obviously, at this stage they will perform poorly and the program will delete the poorest in the list, replacing them with new ones which are made up by cutting up, recombining and mutating selections from the best rules in the list. The rules which are generated in this way can look rather bizarre: Table 2 shows two which emerged from an intermediate stage of one of our runs. Underneath each is its 'score' which measures its performance on a scale from zero to one hundred. For a rule to be of practical use it needs to have a score of over 60 so these two are unlikely to survive subsequent stages.

BEAGLE continues in this way: discarding poor rules; modifying existing rules; mating portions of the better rules to make new rules; testing and repeating. This is why BEAGLE is called 'evolutionary', because it models the way in which, in nature, organisms adapt to their environment. Those that adapt best survive, those that don't perish. It is strikingly different in concept from conventional methods and it consequently brings fresh perspectives to problems such as the one we are discussing. It is difficult to envisage a human expert creating rules such as those in Table 2 and conventional statistical methods certainly would not do so.

There is a price to be paid, of course. All this creation, testing, recreation and refinement of rules requires appreciable computing power and our computer resources five years ago would not have allowed us seriously to contemplate such an approach. Things have changed radically, however, with the advent of cheap processing power.

We started this work on an IBM PC with 256k of memory, later transferring to an Amstrad PC1512 with a 640k memory. A BEAGLE run to derive window/non-window rules based on the entire database of 214 records took approximately one hour using the Amstrad; on the IBM it would have taken about two hours. This run has to be done only once, however, and applying the rules to a new unknown is a fairly simple job. The rules need only be reformulated if the database is updated.

At the end of the run, BEAGLE has a routine which tidies up the surviving rules and attempts to make them appear not too arcane. There is also the useful facility to embed the rules in a PASCAL program for future use.

TABLE 2

EXAMPLES OF BEAGLE RULES FROM AN INTERMEDIATE STAGE OF A RULE INDUCTION RUN ON NON-WINDOWS

((Fe < = NA) and (K > (Fe * 650.00)))

49.73

((Mg < = (K + -0.020000)) : (K < = (Fe + -74.26)))

27.81

The number under each rule is its performance score on a scale from zero to one hundred.

Unlike nearest neighbour and discriminant analysis, BEAGLE rules apply only to either/or classifications so it was necessary to split our problem into steps. First we derived rules for window versus non-window glass, settling for a short list of three rules, then within windows we established three rules for float versus non-float. The non-windows caused us some difficulty because there were three subtypes and because there were relatively few examples of each. We did derive some rules though we would not expect them to be robust.

Comparison of Performance

One method of evaluating the performance of a classification procedure is to try it out on the very data on which it is based. For discriminant analysis and BEAGLE rules this means applying the discriminant functions or rules respectively to each sample in the database in turn and seeing whether the answer is right or wrong. For nearest neighbour the evaluation is done by taking each sample out in turn and comparing it with the rest of the collection.

Both discriminant analysis and BEAGLE provide the means of assessing the uncertainty of a classification and this can also be done, though in a rather ad hoc fashion, with nearest neighbour. However, to keep things simple we decided to adopt a deterministic approach for the comparison: each classification was to be categoric, no hedging was allowed.

The most important step was the window/non-window classification and this was tried out on all 163 windows and 51 non-window samples in the collection. The performance of the three methods is compared in Table 3. BEAGLE was the best performer at spotting window glass samples correctly with an error rate of only 0.6%. All these had the same error rate for spotting non-window glass. It is worth commenting on the association between the different methods. The single window sample which was misclassified by the BEAGLE rules was classified correctly by each of the other methods. In the case of the non-windows there is only one sample which is misclassified by all three methods and that also is the only sample to be misclassified by both BEAGLE rules and discriminant analysis.

Table 4 compares the performance of the three methods at classifying window samples into float and non-float. This was assessed on the basis of 87 float samples and 76 non-float samples. The error rates are much higher. This was expected because this is known to be a difficult task which would not normally be tackled in operational work using RI and elemental analysis alone. BEAGLE's performance is on a par with nearest neighbour and better than that of discriminant analysis. Again the association between the three methods is interesting. Only 5 of the float samples and 6 of the non-float

TABLE 3

COMPARISON OF PERFORMANCE - FULL DATABASE

Classification Task: Is this sample from a window or not?

TRUE TYPE OF SAMPLES	NUMBERS OF INCORRECT ANSWERS		
	BEAGLE	Nearest Neighbour	Discriminant Analysis
Windows (163)	1	5	5
Non-windows (51)	7	7	7

TABLE 4

COMPARISON OF PERFORMANCE - WINDOW GLASS

Classification Task: Is this float glass or not?

TRUE TYPE OF SAMPLES	NUMBERS OF INCORRECT ANSWERS		
	BEAGLE	Nearest Neighbour	Discriminant Analysis
Windows (87)	10	12	21
Non-windows (76)	19	16	22

samples were misclassified by both nearest neighbour and BEAGLE rules: each one of these was also misclassified by discriminant analysis.

Study of the classification within the non-window samples was not very satisfactory because of the small numbers of samples and the results are not reported here, though it is worth saying that BEAGLE's performance was at least on a par with the the two methods.

Overall these results are interesting because they show that BEAGLE performs as well as, if not better than, conventional statistical methods at the chosen tasks. In addition, the fact that BEAGLE rules tend to misclassify different samples confirms that it can provide a fresh insight of existing problem areas. It might be feasible to devise some combination of the schemes which would be more powerful than the individual methods but we haven't explored this prospect.

A Blind Trial

Although it is useful to evaluate a classification scheme by applying it to the collection on which it is based, it is essential for a realistic evaluation that it be tested on fresh unknown samples. For the purpose of this study, a colleague provided ten samples of glass of various types and they were sent to Birmingham for analysis. The data were then subjected to each of the classification procedures 'blind' - the originator of the samples was the only person who knew their types. One of us applied the BEAGLE rules and the other applied the two statistical methods. For each unknown the classification was carried out in two stages:-

(a) Is this sample from a window?

(b) If the answer to (a) is 'yes' then is it float?
If the answer to (a) is 'no' then is it container, tableware or headlamp glass?

Again we adopted a deterministic procedure, giving a yes or a no even in those cases where we knew there was appreciable uncertainty. When the exercise was completed, the types of the samples were disclosed and the three methods were marked. Table 5 shows the results. Note that if the first question is answered incorrectly then the second question is ignored for scoring.

Sample 3 was wrongly classified as a window by all three methods. Discriminant analysis gave two further errors: the main classification of 7 and the subclassification of 10. Apart from 3, nearest neighbour gave one other error: the subclassification of 10. The results reported from BEAGLE rules gave the lowest error rate.

TABLE 5

COMPARISON OF PERFORMANCE - BLIND TRIAL

SAMPLE NUMBER	CORRECT TYPE	BEAGLE	Nearest Neighbour	Discriminant Analysis
1	Non-window	c	c	c
	Headlamp	c	c	c
2	Window	c	c	c
	Float	c	c	c
3	Non-window	w	w	w
	Container			
4	Non-window	c	c	c
	Container	c	c	c
5	Window	c	c	c
	Non-float	c	c	c
6	Window	c	c	c
	Non-float	c	c	c
7	Non-window	c	c	w
	Container	c	c	
8	Window	c	c	c
	Float	c	c	c
9	Window	c	c	c
	Non-float	c	c	c
10	Window	c	c	c
	Non-float	c	w	w

c - Correct Response.
w - Wrong Response.

Discussion

Our work on glass classification is not complete because there are other glass data collections to look at and the results must be scrutinised by colleagues who are established experts in the field. However, we have already gained valuable experience from our rule induction work though we confess to a certain amount of initial scepticism. BEAGLE is inexpensive - less than £100 - and there is a temptation to regard it as a bit of a toy but serious usage demonstrates otherwise. We found a number of quirks in the various routines but overall the package pleased us with its robustness and ease of use.

Forsyth warns that it is possible to do silly things with BEAGLE, the following is a quotation from the manual:-

> There is no magic in BEAGLE, so it cannot find patterns that do not exist. It may be that you are expecting to predict the price of gold in London from the number of motor accidents in New Zealand (so to speak). This example is somewhat far-fetched, but it is all to easy to collect data that is available rather than data that is relevant.

But these remarks apply to any sort of data analysis. It is our impression that, given time for familiarisation, it is reasonable to expect a person who has no knowledge of formal statistics to use BEAGLE intelligently and get useful results. Letting a non-statistician loose on standard statistical packages such as SPSS is a recipe for disaster!

There is room for improvement in BEAGLE but we have found that even in its present form it is a useful weapon to have in the data evaluation armoury. On the task reported here it performed at least as well as the statistical big guns. When it did misclassify this tended to be on different samples from those which were misclassified by the other methods, confirming its originality.

BEAGLE will continue to be used at the Central Research Establishment for this project and for looking at new problems. We can recommend its use to other workers who encounter similar problems.

Acknowledgements

The data for this study were provided by Mr B German of the Home Office Forensic Science Laboratory, Birmingham. The glass samples for the blind trial were prepared by Dr J Locke of the Central Research Establishment. The SPSS discriminant analysis calculations were carried out by Dr P E Cage, also of the Central Research Establishment.

AUTOMATED REQUEST TO THE BRITISH LIBRARY DOCUMENT SUPPLY CENTRE

B COPY

T

17399641

BEAGLE

141

Shelfmark
5088.238300

KBS IN GOVERNMENT
RULE INDUCTION IN FORENSIC SCIENCES
EVETT,I. & SPIEHLER,E.
1987 P.107-18
I0125345

UK FLOOR 4-3
KBS IN GOVERNMENT -CONFERENCE-
D2 S 1987

17399641

INTER LIBRARY LOANS OFFICE
MAIN LIBRARY
UNIVERSITY OF BIRMINGHAM
EDGBASTON
BIRMINGHAM B15 2TT

X
001

WEST M

Return Date:

Request Ref. No.
ND78505 Y

If no other library indicated please return loan to:-
The British Library Document Supply Centre, Boston Spa, Wetherby, West Yorkshire, LS23 7BQ