EXPLOITING A DOMAIN MODEL IN AN EXPERT SPECTRAL ANALYSIS PROGRAM

David R. Barstow

Schlumberger-Doll Research, Ridgefield, Connecticut 06877

ABSTRACT

Gamma ray activation spectra are used by nuclear physicists to identify the elemental composition of unknown substances. Neutron bombardment causes some of the atoms of a sample to change into unstable isotopes, which then decay, emitting gamma radiation at characteristic energies and intensities. By identifying these isotopes, the composition of the original substance can be determined. GAMMA is an expert system for performing this interpretation task. It has a detailed model of its domain and can exploit this model for a variety of purposes, including ratings for individual isotopes and elements, ratings based on multiple spectra, complete interpretations, and even calibration. GAMMA's performance is generally quite good when compared with human performance.

1. INTRODUCTION

Gamma ray spectra are commonly used by nuclear physicists to identify the elemental composition of a substance. One kind of gamma ray spectrum (an "activation spectrum") is produced by bombarding a sample of the substance with neutrons. This causes certain changes in some of the atoms in the sample, many of which result in unstable isotopes that then begin to decay. As they decay, the unstable isotopes emit gamma rays at characteristic energies and intensities. By measuring these, the unstable isotopes (and from these, the elements of the original sample) can be identified. For example, Figure 1 shows such a spectrum, with peaks identified and labeled by a physicist. In this case, the peaks were produced by emissions from the isotopes Na-24, Cl-37, and S-37; the original substance was a sample of salt.

An expert system, called GAMMA, has been developed to perform this task, and GAMMA's performance compares well with human interpreters. The basic strategy employed in developing GAMMA was to develop a detailed model of the domain and then to exploit this model for a variety of tasks and situations. Early work on GAMMA was discussed in another paper[1]; in this paper, recent progress will be discussed.

2. The Domain Model

GAMMA's domain model was described in detail in the earlier paper and will only be summarized here. Basically, the process that produces gamma ray activation spectra can be seen at six different levels as follows:

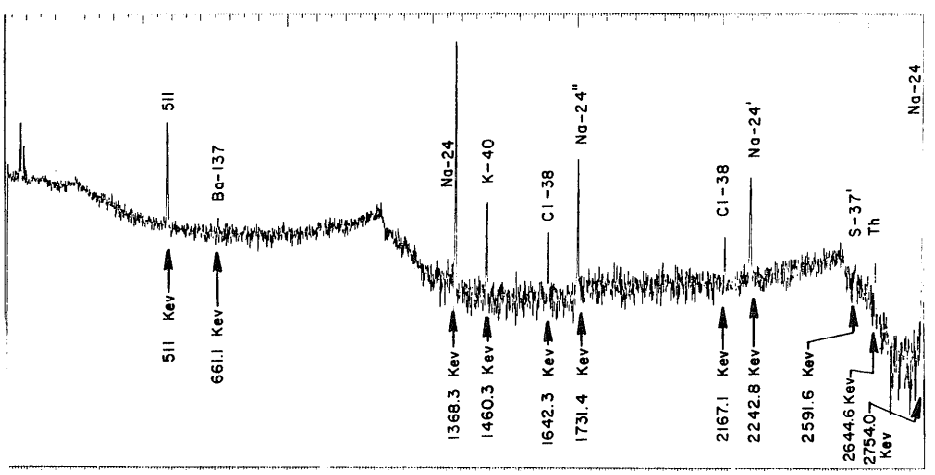

Figure 1: Gamma Ray Activation Spectrum

(1) elements in the original sample

(2) isotopes in the original sample

(3) isotopes after bombardment by neutrons

(4) decays of unstable isotopes

(5) gamma ray emissions during decay

(6) gamma ray detections during decay

Level 6 represents the actual spectrum and level 1 represents the ultimate goal of the interpretation. Level 3 is a convenient intermediate level used by most practicing nuclear physicists.

GAMMA's use of this domain model involves considering hypotheses at each of the levels. A hypothesis consists of a set of triples, each triple consisting of an object appropriate to the level (e.g., naturally occurring isotope for level 2; gamma ray emissions at a particular energy for level 5), an estimated concentration for the object (e.g., the number of decays of a given unstable isotope for level 4), and a label which encodes the path from level 1 to the triple (e.g., "NA-23/NG/NA-24" denoting that the unstable isotope Na-24 was produced when the naturally occurring isotope Na-23 underwent an N-γ transition).

Relationships between levels can be expressed in terms of several formulae that have been derived from both theoretical considerations and empirical observations. These formulae involve such parameters as the likelihood of particular isotopic transitions during neutron bombardment, the half-lives of unstable isotopes, and the characteristic gamma ray energies and intensities for different isotopes. Nuclear physicists consult published reports when they need such information; the data from one such source[2] has been converted into a LISP data base for GAMMA's use. Further details of the formulae and data base are available elsewhere[1].

In GAMMA's case, the formulae are all used predictively; that is, given an hypothesis at one level, the appropriate formula can be used to predict hypotheses at the next level down. By chaining such predictions together, GAMMA can go from hypothetical interpretations at levels 1 or 3 down to predicted gamma ray detections that can be compared against spectral data.

3. Applications of the Domain Model

The accuracy with which predictions can be made and the high resolution of this particular detector enable GAMMA to exploit the domain model in a variety of tasks and situations. Some of these were discussed earlier[1], and will be mentioned here only briefly.

3.1. Isotopic and Elemental Ratings

GAMMA was first used to "rate" the likelihood that any particular unstable isotope was present after neutron bombardment. This was done by hypothesizing decays of that isotope at level 4, predicting detections at level 6, and using an evaluation function to compare the predictions with the spectral data. The evaluation function was designed to take positive and negative evidence into account, allowing both for background radiation and for noise and errors in the prediction and detection processes. The peaks were individually rated for both energy and intensity, and the final rating was the average of the individual ratings, weighted by intensity (i.e., stronger peaks were more important). When a predicted peak had a corresponding detected peak, a positive rating was given; when no peak was detected, a negative rating was assessed, unless the predicted intensity was low enough that the peak could have been obscured by background radiation. Noise and errors were taken into account by using what we call the trapezoidal rule. For example, the trapezoidal rule for peak energies is shown in Figure 2. If a peak was predicted at energy E, then a detected peak within the range $(E-\delta_1, E+\delta_1)$ was considered a perfect match, peaks outside the range $(E-\delta_2, E+\delta_2)$ were not matched at all, and peaks in the ranges $(E-\delta_2, E-\delta_1)$ and $(E+\delta_1, E+\delta_2)$ were scaled to provide a continuous function.[2] Such trapezoidal rules were used throughout GAMMA's evaluation function, and has proved quite adequate. GAMMA's performance at the isotopic rating task was moderately good compared with that of human experts: although it gave high ratings to isotopes identified by experts, it also occasionally gave such ratings to isotopes considered implausible by the experts.

GAMMA's second task was to do a similar rating for elements in the original sample (i.e., hypotheses at level 1). The same predict-and-match technique was used, and GAMMA's performance was again moderately good, although not quite as good as in the isotopic case: fewer implausible elements were rated but some elements identified by the human experts received low ratings. This was due largely to certain simplifying assumptions in the formulae relating levels 2 through 4. Further details of GAMMA's rating scheme are given elsewhere[1].

Recently, GAMMA's repertoire has been expanded to include several other tasks, and its performance seems to have improved with age.

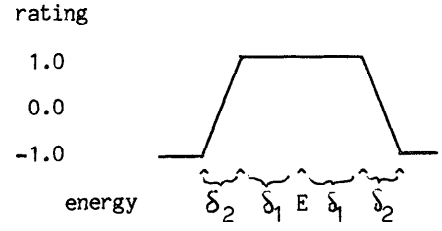

Figure 2: Trapezoidal Rule for Peak Energies

3.2. Ratings for Multiple Spectra

GAMMA's next major task was to do similar ratings for individual isotopes and elements, but to do so on the basis of multiple spectra: in a typical experimental situation, not one but several spectra are recorded, each for a different time interval. Generally, the first few spectra are for comparatively short time periods (10 to 30 seconds), and the later spectra may be for periods as long as several hours. The primary advantage of multiple spectra is that they permit greater use of half-life information: unstable isotopes with short half-lives will appear on the earlier spectra but not on the later ones; isotopes with longer half-lives emit gamma rays at roughly constant rates, so they appear most distinctly on the later spectra (for which the detection time is longer).

The technique used by GAMMA is to find the isotopic (or elemental) ratings for the individual spectra, use these to hypothesize an initial concentration for the isotope, redo the predictions based on this concentration, and finally combine the ratings for these predictions into a single overall rating. The hypothetical concentration for an isotope is determined by considering all spectra for which the isotope's rating is sufficiently high, taking the concentrations (byproducts of the original prediction-and-match rating) that agree sufficiently well (within one order of magnitude), and then computing the average. This technique is designed to ignore those results which, for any of several reasons, deviate from the norm and in practice seems to work quite well. Given this hypothesized concentration, the prediction-and-match rating is again computed for each individual spectrum. These ratings are then averaged to determine the overall "multiple spectra" rating for the isotope.

In our first attempt to average these ratings, we weighted them by the total predicted intensity for a spectrum, as was done for ratings within individual spectra. But this seemed to attach too much weight to spectra with high predicted intensities, so on our second attempt, we took the simple average of the ratings for all spectra for which the evidence was significant (either positive or negative), and the results were much better. (Interestingly, the first symptom of this problem was due to an INTERLISP error: under certain circumstances, INTERLISP ignores floating point overflow and underflow, thereby producing a very large number when multiplying two very small ones. With simple averaging, such isolated erroneous computations no longer have much overall effect. In fact, we now take this as a maxim: no numeric rating scheme should depend too heavily on any single data point.)

GAMMA's performance on multiple spectra is generally much better than on individual spectra, primarily because of the value of half-life information. GAMMA's ratings generally compare well with those of human experts, and implausible isotopes (or elements) are only rarely given high ratings.

3.3. Producing a Complete Interpretation

The major problem with the tasks described so far is that the ratings are given to isotopes and elements as if they were totally independent of each other. The fact that the same peak may be caused by emissions from two different isotopes does not detract from the rating of either one. The ultimate interpretation of spectral data should not be ratings for individual elements, but rather a set of elements (and concentrations) which, taken together, explain the data well. A first pass at coming up with such a complete interpretation might be to take all and only those elements with sufficiently high ratings, but that does not take into account the interaction between the elements, and is simply inadequate. GAMMA's solution to this problem is essentially a hill-climbing algorithm designed to maximize an "interpretation measure".

For this algorithm, a complete interpretation is defined to be a set of <element, concentration> pairs, and a mapping of detected peaks to sets of labels. (The labels describe the path from one of the <element, concentration> pairs to the detected peak. Under this definition, the same detected peak may have several different labels, a situation which actually occurs in the spectra under consideration.)

The interpretation measure that GAMMA currently uses is based on two different considerations. First, the individual spectra are rated in terms of (1) how many peaks have no labels (i.e., are there peaks which are not explained by the interpretation?), (2) how many labels have no peaks (i.e., are there predictions which do not appear in the detected spectra?), and (3) how well the peaks and associated labels match (i.e., do the energies and intensities of the detected peaks match well with the energies and intensities predicted for the associated labels?). The second consideration is that the relative concentrations of the elements be plausible. This is used only as negative evidence: if the concentration of an element is high (relative to the concentrations of the other elements), but the rating for that element is low, then the interpretation is suspect, since the detector and model can be expected to be quite accurate for relatively pure substances; if the concentration is below a certain threshold, then the interpretation is also suspect, since the detector simply cannot be expected to find elements in such small concentrations.

The task is thus to find the set of <element, concentration> pairs that maximizes this measure. GAMMA uses the following hill-climbing algorithm:

```
INTERPRETATION := {};
CANDIDATES := {<element, concentration> |
                  rating is above a threshold};
consider all interpretations formed by moving
    one element from CANDIDATES to INTERPRETATION
    or from INTERPRETATION to CANDIDATES;
if no such interpretation increases the measure
    then quit
    else select that which maximizes the measure;
         update INTERPRETATION and CANDIDATES;
         repeat.
```

While we have no theoretical basis for claiming that this algorithm does, indeed, find the subset of candidates with maximal measure, our experience indicates that it performs very well, and the interpretations that GAMMA produces are quite good.

3.4. Calibration

GAMMA's final task is to calibrate the spectra. Up to this point, it has been assumed that the input spectra have already been calibrated (spectral channels have been associated with gamma ray energies), and this is a task which has hitherto been performed by physicists before the data are given to GAMMA. We have not yet completed the attempts at solving this problem, but our first results are encouraging: we have developed a technique for recalibrating a spectrum more precisely, given an initial approximate calibration. The basic technique is to use a set of good calibration isotopes (as identified *a priori* by physicists) and look for any which satisfy two criteria: (1) there is exactly one match within a given range of the initial calibration; (2) that match gets a fairly high rating. A linear least squares fit of these points gives the recalibration. Experience with this approach is quite good, and it is currently used to correct for calibration differences among the individual spectra within a set of spectra.

Our plan for finding the initial calibration is to take a sum spectrum for the entire set (thereby increasing the data/noise ratio) and apply a similar strategy: find any "calibration" isotope which has exactly one match anywhere on the sum spectrum and for which the match rating is quite high. We hope to know soon whether this approach will succeed.

4. Shallow and Deep Domain Models

GAMMA's success is due largely to its use of a relatively detailed model of its domain. This may be compared with systems such as MYCIN[3] whose success is due largely to shallow models that encode, in a sense, an expert's compiled version of a more detailed model that the expert may or may not know explicitly. In comparing these two approaches, several observations can be made. First, while a deep model can be put to great use (as it was in GAMMA), there are several circumstances in which a shallow domain model is necessary: (1) a deep model doesn't exist (e.g., there is no computational theory of the entire human body); and (2) a deep model is not computationally feasible (e.g., one cannot hope to do weather prediction based on the fundamental properties of gases). Second, although shallow models will often suffice, it seems likely that future expert systems based on shallow models will require access to deep models for difficult cases. Third, "deep" and "shallow" are obviously relative terms: GAMMA's deep model is shallow when viewed as a model of subatomic physics. The relationship between deep and shallow models seems to be an important topic for future work on expert systems.

5. References

[1] Barstow, D.R.
Knowledge engineering in nuclear physics.
In Sixth International Joint Conference on Artificial Intelligence, pages 34-36.
Stanford Computer Science Department, 1979

[2] Erdmann, G.
Neutron Activation Tables.
Verlag Chemie, New York, 1976 .

[3] Shortliffe, E.H.
MYCIN: Computer-Based Medical Consultations.
American Elsevier, New York, 1975 .