

KNOWLEDGE ENGINEERING IN NUCLEAR PHYSICS

David R. Barstow
Department of Computer Science
Yale University
P.O. Box 2158 Yale Station
New Haven, CT 06520

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 the sample to change into unstable isotopes, which then decay, emitting gamma radiation at characteristic energies and intensities. By identifying the unstable isotopes, the composition of the original substance can be determined. Since the performance of such analysis relies on large amounts of various kinds of knowledge, the task seems appropriate for the techniques of knowledge engineering. An experimental system, GAMMA, has been developed, based on the generate-and-test paradigm. GAMMA's performance has been good enough that it is currently in use by practicing nuclear physicists.

1. INTRODUCTION

Gamma ray activation spectra are produced by bombarding a sample with neutrons, thereby producing unstable isotopes which begin to decay. While decaying, these isotopes emit gamma rays at characteristic energies and intensities. By measuring the gamma rays emitted by the sample after bombardment, the unstable isotopes (and from these, the elements of the original sample) can be identified. Fig. 1 shows the spectrum produced after bombarding a sample with neutrons. Some of the peaks are labeled by their energies. This is considered to be a high resolution spectrum, since the energies can be identified quite accurately. Fig. 1 also shows the isotopic interpretation made by a nuclear physicist. The peaks are labeled by the isotope which emitted those gamma rays during decay. The elemental interpretation attributes the Na-24 to sodium, the Cl-38 and S-37 to chlorine, the K-40 and Th to natural radiation, and the Ba-137 to an impurity.

The task of interpreting gamma ray activation spectra requires considerable knowledge about nuclear physics, suggesting that the techniques of knowledge engineering may be usefully

The work reported herein was accomplished by the author while serving as a consultant to the Computer Intelligence Program at Schlumberger-Doll Research Center. A detailed version of this paper is available from: Dr. W. Frawley, Schlumberger-Doll Research Center, Old Quarry Road, Ridgefield, CT 06877.

applied. Some of this knowledge has been codified into a machine-usable data base, and an experimental program, called GAMMA, has been implemented for using this data base within a generate-and-test paradigm. Preliminary results with GAMMA have been good enough that even the experimental version has already been used for cross-checking analyses by several practicing nuclear physicists.

2. KNOWLEDGE BASE FOR GAMMA RAY ANALYSIS

The process that produces gamma ray spectra can be seen at six different levels as follows:

- (1) elements in original sample
- (2) isotopes in original sample
- (3) isotopes after bombardment
- (4) decays
- (5) emissions during decay
- (6) detections during decay

The relationships between the levels can be understood in terms of the different kinds of knowledge that play a role. The relationship between levels (1) and (2) involves the relative concentrations of the naturally occurring isotopes of each element. The relationship between levels (2) and (3) is more complex. Under neutron bombardment, a given isotope can go through any of four basic transitions, and the frequency with which the different transitions may occur depends on the particular isotopes involved, on the device used to produce

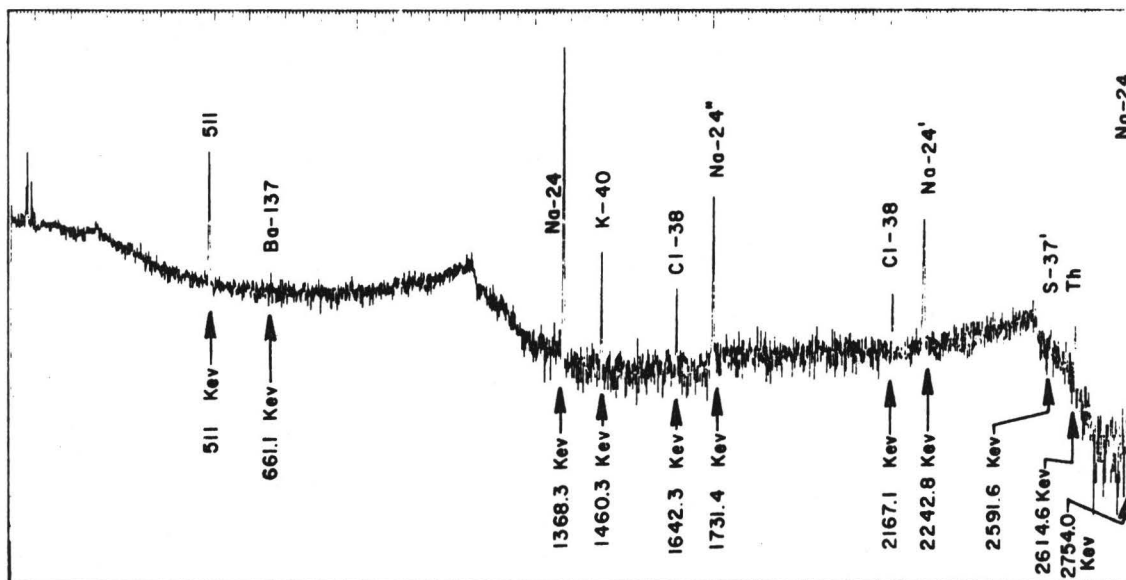


Figure 1.

the neutrons, and on the length of time during which the sample is exposed to neutrons. The principal factors relating levels (3) and (4) are the half lives of the unstable isotopes and the time during which the detector is exposed to the sample after bombardment. The relationship between levels (4) and (5) is relatively simple: every unstable isotope has a characteristic set of gamma rays (with relative intensities) emitted during decay. The relationship between levels (5) and (6) depends on the detector: any given detector will detect some fraction of the emitted gamma rays. This detector efficiency may depend on the energy of the gamma rays. A second factor, the presence of escape peaks (e.g., those labeled Na-24' and Na-24'' in Fig. 1), also depends on the energy of the main peak.

Much of the knowledge discussed in this section is available to nuclear physicists in the form of books and articles; we have converted the data from one of these [1] into a machine-usable LISP data base.

3. PARADIGMS FOR KNOWLEDGE ENGINEERING

Based on the discussion of the preceding section, several implementation paradigms can be imagined. Generate-and-test could be based on a list of elements. For each element, one could progress from level (1) to level (6), predicting each level from the previous one. The predicted level (6) peaks could then be matched against the peaks of the spectrum to accept or reject the original element. Backward-chaining would involve progressing through the levels in the same order, but in some cases it might be unnecessary to go all the way to level (6) for

all subgoals, because rejection of one subgoal makes the others irrelevant. In addition, the list of elements would only be implicit in the list of rules relating levels (1) and (2). In a forward-chaining (or data-driven) paradigm, one could work upwards from the spectral peaks at level (6) to level (1), using data (or hypotheses) at each level as evidence for hypotheses at the next. Finally, one could imagine some kind of mixed-chaining, using individual peaks in the spectrum to suggest candidates to be evaluated further. (This is roughly the strategy used by nuclear physicists.) We plan eventually to experiment with all of these paradigms.

4. GAMMA

Our first experimental system, GAMMA, uses the generate-and-test paradigm. GAMMA has routines to predict from hypotheses at one level to hypotheses at the next level. Chaining together predictions from a single element at level (1) to level (6) gives a predicted pattern of peaks, which can then be matched against the peaks of the spectrum to determine whether or not the element was present in the original sample. (Note that this technique only works when the hypotheses (i.e., elements) at the top level can be considered relatively independently.)

Prediction from level (1) to level (2) is quite simple, since the relative abundances of the naturally-occurring isotopes of an element are stored in GAMMA's data base. Prediction from level (2) to level (4) is collapsed into one step, from isotopes in the sample to decays of unstable isotopes after neutron bombardment. We

are currently using an analytically determined formula which incorporates an estimated cross-section for each possible transition, the half-life of the resulting isotope, and the three relevant time periods (bombardment, detection, and the delay between them). Prediction from level (4) to level (5), from decays of single isotopes to emitted gamma rays, is quite simple, since for most isotopes the characteristic gamma ray energies and intensities have been measured and this list is stored in GAMMA's data base. The final prediction step, from level (5) to level (6), is based on two experimentally determined functions relating detector efficiency to a peak's energy.

Due to noise and error, the predicted peaks may not occur in the spectrum at precisely the predicted energies and intensities. To determine whether or not a predicted set of peaks actually occurs, GAMMA currently uses a simple rating scheme. The match value of a set of peaks is the average of the match values of the individual peaks, weighted by intensity of the predicted peak (i.e., predicted peaks of high intensity are more important). For single peaks, there are two cases. (1) If there is no spectrum peak within a window of the predicted peak, and the predicted intensity is larger than the background intensity, a negative rating (based on the ratio of the intensities) is assigned. (2) If there is a spectrum peak within a window, the rating incorporates an energy rating (based on the difference between the predicted and detected energies) and an intensity rating (based on the ratio of the intensities and whether the detected is less than the predicted intensity).

This scheme is intended to make use of both positive and negative evidence, allowing for the imprecision of the prediction and detection processes and for the background radiation. While it is based largely on guesses by nuclear physicists, we have found it to be quite satisfactory. In fact, the difference between the presence and absence of an unstable isotope in such high resolution spectra seems to be great enough that almost any "reasonable" matching procedure would work well, as long as it does not assume too much precision on the part of the predictor and detector.

Let us reconsider the spectrum in Fig. 1 from GAMMA's viewpoint. The statistically significant peak energies and intensities are computed by a preprocessor developed by the human experts. For each element, GAMMA predicts a set of peak energies and intensities, and then matches these against the spectrum peak list. The elements are then sorted by their match values. From this list, the spectrum peaks are

labeled by all elements which matched the peak and whose match rating was above a certain threshold. The resulting interpretation is substantially the same as that of the human nuclear physicists given earlier. The only differences are that GAMMA found no elemental source for the Ba-137 peak (which the physicist attributed to an impurity), and that GAMMA labeled the Th peak as Po-212 (which is one isotope in the Thorium natural radiation chain). For the curious, the substance being analyzed is a sample of Alaskan salt water.

5. DISCUSSION

We selected the generate-and-test paradigm for our first implementation because it is relatively simple to implement. In fact, the major effort involved in building GAMMA was the conversion of the data base from human to machine-usable form. This first version has been fairly successful, and is, in fact, being used as an aid by nuclear physicists. While they are not yet ready to blindly accept GAMMA's interpretations, they have found GAMMA to be useful in cross-checking their own analyses. The fact that this version succeeded fairly well is due primarily to several features of the domain: an exhaustive generator is available, the prediction process is fairly simple, and the high resolution of the spectra permits the use of a relatively imprecise matching process.

In the future, we plan to pursue two directions. First, we hope to improve the current version of GAMMA by incorporating the use of multiple spectra (which would permit more use of half-life information) and re-calibration on the basis of an initial interpretation. Second, we plan to experiment with several other paradigms, including backward-chaining, forward-chaining, and mixed-chaining. In so doing, we hope to learn more about the relative merits of these paradigms, and what domain characteristics are important in selecting one over another.

ACKNOWLEDGEMENTS

Those at SDRC have been of invaluable assistance: nuclear physicists A. Becker and J. Schweitzer; W. Frawley, H. Austin and P. Pruchnik of Computer Intelligence; K. O'Brien and P. Martinich of the office staff. K. Pitman of MIT-LCS designed and built the user interface for GAMMA. R. Davis of MIT-AI and M. Stefik of SU provided helpful comments on earlier drafts.

REFERENCES

[1] Erdtmann, G. Neutron Activation Tables, Verlag Chemie, New York, 1976.