Plasma X-ray Spectra Analysis Using Genetic Algorithms

Igor E. Golovkin
Department of Physics
University of Nevada Reno
Reno, NV 89557
golovkin@physics.unr.edu

Roberto C. Mancini
Department of Physics
University of Nevada Reno
Reno, NV 89557

Sushil J. Louis
Department of Computer Science
University of Nevada Reno
Reno, NV 89557
sushil@cs.unr.edu

Abstract

X-ray spectroscopic analysis is a powerful tool for plasma diagnostics. We use genetic algorithms to automatically analyze experimental X-ray line spectra and discuss a particular implementation of the genetic algorithm suitable for our problem. Since spectroscopic analysis may be computationally intensive, we also investigate the use of case injected genetic algorithms for quicker analysis of several similar (time resolved) spectra. Preliminary results are promising and genetic algorithms seem to provide a reliable and robust approach for automated analysis of X-ray line spectra.

1 INTRODUCTION

X-ray spectroscopic analysis is a widely used method for hot dense plasma diagnostics (Griem, 1992). Spectroscopy diagnostics is important for astrophysical as well as laboratory plasmas. In this paper we discuss spectroscopic analysis of experimental X-ray line spectra obtained during Inertial Confinement Fusion (ICF) experiments. The idea is the following: given a physics model for the calculation of X-ray line spectra that depend on plasma parameters (e.g., temperature and density) a genetic algorithm based procedure automatically looks for the best synthetic spectra fit to experimental data. In turn, determination of temperature and density during ICF implosion experiments is important to provide insight into studying the possibility of controlled thermonuclear fusion.

In many cases observation of radiation that comes from plasmas is the only source of information available. The analysis of recorded spectral intensity distribution of the radiation (experimental spectra) usually proceeds as follows:

- A theoretical synthetic spectrum is computed depending on plasma parameters that define the intensity distribution of radiation coming from the object,
- Experimental spectrum and theoretical fit are compared.
- Based on the quality of the comparison a new set of plasma parameters and synthetic spectrum are computed.
- These steps are iterated until the best theoretical fit to the experiment is obtained.

If the theoretical spectrum matches experimental data, we assume that the parameters used to construct the synthetic spectrum are representative of the state of the plasma during the formation of the X-ray line spectra. In general real spectra are very complex and the intensity distribution depends quite non-linearly on the fitting parameters (in our case: temperature, density, and shift). These parameters may vary within a known range, and the problem is to find the best combination within that range. The complexity of the parameters' functional dependence makes it difficult for conventional search algorithms to find good solutions. Thus, looking for the best fit to experimental data may be a lengthy and laborious procedure involving much twiddling of parameter values. In this paper, we investigate the application of Genetic Algorithms (GAs) for this task (Holland, 1975; Goldberg, 1989). The GA does not perform a simple random walk but does a selective exploration of the search space. Its robustness in finding a solution in a poorly understood search space was our primary motivation in choosing it as our search engine for fitting experimental X-ray spectra.

Another problem investigated in this paper concerns time resolved spectra that provide important information about evolution of plasmas. Analysis of such spectra allows us to determine plasma temperature and density as a function of time. Usually there are no abrupt changes of plasma parameters and therefore subsequent spectra are similar. Furthermore, spectra may be simultaneously collected by several instruments and need to be analyzed. In other words, time resolved and multi-instrument spectra lead to sequences of similar spectral analysis problems. It makes little sense using a randomly initialized genetic algorithm for each time interval or for each instrument when information from previous search attempts on similar problems is available. Case Injected Genetic AlgoRithms (CIGARs) were designed to learn to solve such similar problems quickly (Louis and Johnson, 1997). We therefore use CIGARs to accelerate spectroscopic analysis and/or produce better theoretical fits to sets of similar experimental spectra.

CIGARs combine genetic algorithms with a case-based memory to improve performance at related tasks. Typically, a genetic algorithm randomly initializes its starting population so that the GA can proceed from an unbiased sample of the search space. Instead, periodically injecting a genetic algorithm's population with relevant solutions or partial solutions to similar previously solved problems can provide information (a search bias) that reduces the time taken to find a quality solution. In CIGAR, the data-base, or case-base, of problems and their solutions supplies the genetic problem solver with a long term memory. The system does not require a case-base to start with and can bootstrap itself by learning new cases from the genetic algorithm's attempts at solving a problem. Figure 1 shows a conceptual view of CIGAR. When confronted with a problem, we **periodically inject** a small number of *solutions* similar to the current best member of the GA population (closest to the best) into the *current* population, replacing the worst members. The GA continues searching with this combined population. During a GA search, whenever the fitness of the best individual in the population increases, the new best individual is stored in the case-base. Reusing old solutions has been a traditional performance improvement procedure. Our work differs in that we attack a set of tasks, store and reuse intermediate candidate solutions which results in better performance, and do not depend on the existence of a problem similarity measure avoiding indexing problems common to case-based systems.

What happens if our similarity measure is noisy and/or leads to unsuitable retrieved solutions? By definition, unsuitable solutions will have low fitness and will quickly be eliminated from the GA's population. CIGAR may suffer from a slight performance hit but will not break or fail – the genetic search component will continue making progress towards a solution. CIGAR is robust. Note that we can choose schemes other than injecting the closest to the best; schemes such as furthest from the worst and probabilistic versions of both have proven effective.

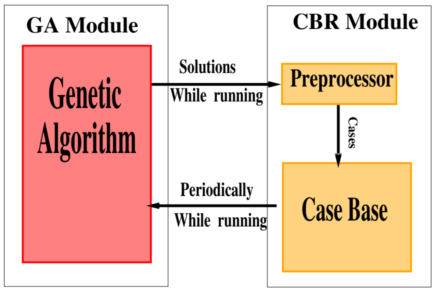

Figure 1: Conceptual view of CIGAR.

We performed several calculations using a simplified physical model (in order to reduce computing time) with the goal of studying different crossover, mutation, and selection operators as well as learning strategies and their effect on convergence. Computing times vary from 2-3 seconds per evaluation for the simple model to several minutes per evaluation for complex models on an SGI Power Challenge machine. The results are quite promising and will allow us to apply the techniques proposed in this paper (GAs and CIGAR) to more complex problems that include parameterization of the temperature and density gradients for non-uniform plasmas.

2 METHODOLOGY

Computing a theoretical synthetic spectrum of the radiation coming from hot dense plasma is a non-trivial problem. For our purposes we used the LAMBDA model and code (Golovkin and Mancini, 1999) for uniform, optically thin plasmas. This model calculates synthetic spectra in an one-dimensional plane-parallel slab plasma approximation. As boundary conditions we use absence of incident radiation on either side of the plasma slab. Atomic processes included in the model are electron collisional excitation and de-excitation, electron collisional ionization and recombination, autoionization and electron capture, spontaneous radiative decay, stimulated emission, photoexcitation, and radiative recombination. Atomic rates, cross sections and energy level structure were calculated with the Los Alamos National Laboratory atomic structure and scattering codes. Collisional rates were computed by integrating cross sections with a Maxwellian electron distribution function. Intrinsic line profiles in the model are Stark-broadened line shapes calculated in the standard Stark broadening theory approximation using a multielectron radiator line profile code that takes into account the effects due to the microfields of the plasma ions and electrons. The effect of the ions was calculated in the static ion approximation (except for the resonance line that includes ion dynamics effects), while that of the electrons using a quantum-mechanical second-order relaxation theory. The ion microfield distribution function was calculated using the APEX model assuming equal ion and electron temperatures. No line shifts have been considered in these line profiles. All these data are input to our model that self-consistently solves a set of collisional-radiative atomic kinetics equations and the radiation transport equation. This is important for plasmas that are not in local thermal equilibrium. We utilize an iterative procedure to calculate atomic population and radiation field.

The problem can be tremendously simplified if one assumes that the plasma is optically thin. This assumption allows us to uncouple atomic kinetics and radiation transport equations and significantly decrease execution time since radiation dependent rates that involve space, angle, and frequency integration no longer need to be computed.

These approximations allowed us to significantly reduce the time required for computing synthetic spectra while keeping the flavor of the model intact. This sacrifice of several physical processes was necessary to allow us to perform a series of numerical experiments to study the effects of different GA operators and parameters and to evaluate the genetic algorithm's effectiveness as a spectra-fitting tool. We plan on using the more computationally intensive but physically accurate plasma models after this feasibility study and GA tuning.

We chose the He_β and associated Li-like satellites composite spectral feature as a characteristic example of X-ray line spectra widely used for spectroscopy diagnostics. Figure 2 displays an experimental spectrum of this feature and a theoretical fit. Also indicated in the figure are the density, temperature and shift obtained from the analysis. The experimental spectrum was recorded during a laser-driven implosion performed at the Laboratory for Laser Energetics, University of Rochester (Haynes and et. al., 1996).

The distance between experimental and theoretical

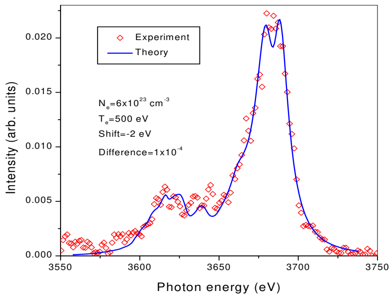

Figure 2: Experimental X-ray spectrum and theoretical fit.

spectra is defined as the sum of squared differences of intensities of the spectra evaluated at the experimental photon energy points (EPEP). That is,

$$\text{distance} = \sum_{\text{EPEP}} (I_{exp} - I_{theor})^2.$$

We use this formula since the photon energy points are the same for all experimental spectra in our problem. In the more general case, when the number of experimental points varies from one spectrum to another, we would use mean squared error or other ways to measure the distance. The intensity of the theoretical spectrum needs to be adjusted automatically to minimize the difference. While temperature and density define a shape of the theoretical spectrum, the shift helps to correct for plasma effects as well as uncertainty in the experimental data. Since we want to minimize the sum squared distance (error) and genetic algorithms always maximize fitness we define fitness as

$$\text{fitness} = \frac{1}{\text{distance}}$$

and therefore convert the minimization of distance into the maximization of fitness. Figure 3 shows a typical evaluation function.

Temperature, density, and shift vary within a known range. For each of the parameters we use a binary 5-bit encoding that results in a 15-bit chromosome. This encoding provides sufficient precision for plasma diagnostic purposes. Figure 4 illustrates the encoding and mapping of our parameters.

We expect our proposed method to work for more complete physical models and therefore implemented an elitist selection scheme modeled after the CHC algorithm with linear scaling (Louis and Rawlins, 1991) that is claimed to converge quickly. A combination of conservative selection, highly destructive crossover and mutation operators, and fitness scaling makes the algorithm quite powerful and robust (Eshelman, 1991).

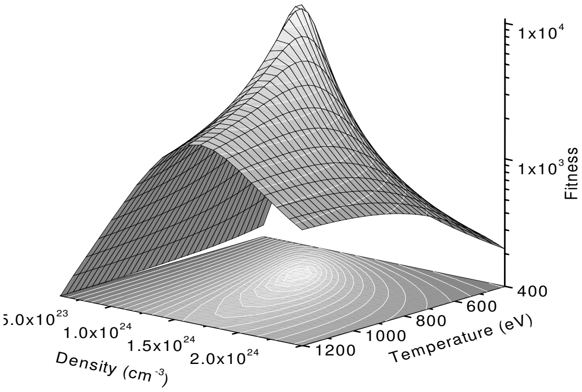

Figure 3: Fitness as a function of two fitting parameters: Temperature and Density

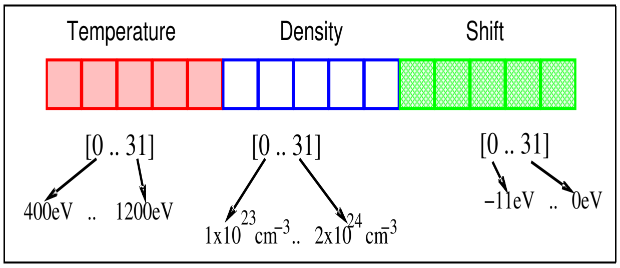

Figure 4: Encoding of temperature density and shift

We use a case injected genetic algorithm on seven (7) time resolved spectra. There are two performance criteria for evaluating CIGAR's efficacy: Quality of solution and time taken to find the highest quality solution. An increase in the quality of solution with number of problems solved and/**or** a decrease in time needed to produce a high quality solution will support the validity of this approach.

The next section presents our results with GAs applied to single spectra and CIGAR applied to multiple time-resolved spectra.

3 RESULTS

We performed a series of experiments to adjust GA parameters so that we can reliably find good quality solution quickly. Analysis of the results from these experiments with spectra taken from plasmas with different parameters convinced us that the GA reliably converges to a physically meaningful result when using the parameters in Table 1.

For each experimental spectrum we run the GA ten (10) times with different initial random seeds. Figure 5 shows maximum, average, and minimum fitness averaged over ten runs as a function of generation number

Table 1: GA parameters that lead to fast convergence

Chromosome length	15
Population size	40
Maximum number of generations	40
Crossover probability	0.95
Mutation probability	0.05
Fitness scaling factor	1.2

for a characteristic experimental spectrum.

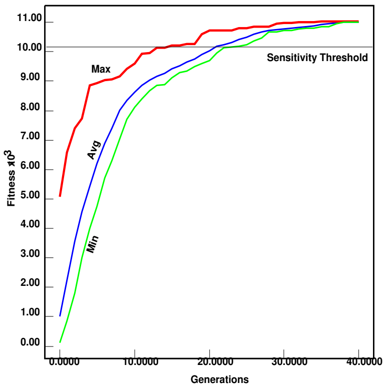

Figure 5: Maximum, average, and minimum fitness vs. number of generations

All the results are acceptable even though not all of the runs produced maximum fitness. Uncertainty in experimental data does not allow claiming very high precision in temperature and density diagnostics with the results that we have the sensitivity threshold indicated in the figure. Table 2 tabulates GA performance for three spectra.

Table 2: Performance on three sample spectra

	Spec1	Spec2	Spec3
Number of runs	10	10	10
Best fitness found	7143	11322	8809
Worst fitness found	3303	90505	8507
Acceptable results	9/10	10/10	10/10
Avg. convergence speed	520	480	200

These results show the robustness of the algorithm. Almost all runs had meaningful results and the average number of evaluations required to get a good solution (450) is small compared to the size of the search space (2^{15}).

We next applied CIGAR and GA to seven time resolved spectra. Note that the fitness of the solution for each spectrum depends on the quality of the data as well as the intensity of the radiation. Since intensity may change significantly during the experiment, we normalized our experimental spectra to assure comparable fitnesses for all the cases. When applied to the first spectrum we expect no performance difference between the GA and CIGAR since there is no case-base and both start from randomly initialized populations. Subsequent spectra should show a performance difference with the gap widening as more spectra are analyzed. Figure 6 compares convergence behavior on the last problem attempted and we can see that CIGAR converges more quickly.

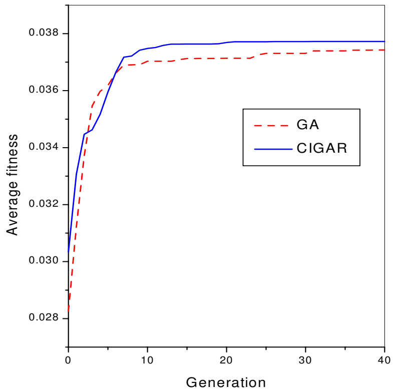

Figure 6: Avg. maximum fitness vs. number iterations on problem 7

There are two distinct phases in plasma behavior over the time period under consideration. During the first phase the plasma implodes and temperature grows. The first three problems correspond to this phase while the fourth problem corresponds to the temperature peak. From this point, the plasma begins to cool accompanied by a substantial change in spectra. We should therefore also expect two distinct phases in CIGAR's behavior.

We ran both the GA and CIGAR ten (10) times with different random seeds and use the GA parameters in Table 1 for both. CIGAR replaces fifteen percent (15%) of the population with individuals from the case base every four (4) generations. We replace the worst members in the population and choose individuals in the case base that are closest in hamming distance to the best individual in the current population. We stop injecting after 20 generations. Increasing the injection percentage beyond a certain limit tends to degrade performance and usually leads to premature convergence. In our case 15% worked well although we did not have time for extensive experimentation. It takes approximately five hours to attempt all seven problems once. Injecting after about half the maximum number of generations does not significantly affect performance since the population diversity is low and injected individuals tend to be repeated.

The injection period of 4 generations was chosen to be less than the minimum takeover time of an individual. With our elitist selection scheme an individual could take over the population in $O(log_2(\text{population size}))$ generations. This period balances two parameters: 1) the time needed by injected individuals to contribute their genetic material without needing to compete with newly injected individuals, and 2) the minimum time needed before an (injected) individual can takeover the population.

Figure 7 compares the quality of solutions produced by the GA to the quality of those produced by CIGAR. The figure plots the number of spectra analyzed on the x-axis versus the average, over ten runs, of the maximum fitness on the y-axis. We can see that there is no significant difference in quality. Fitness decreases from problem 1 to problem 7 because the noise level increases during the experiment.

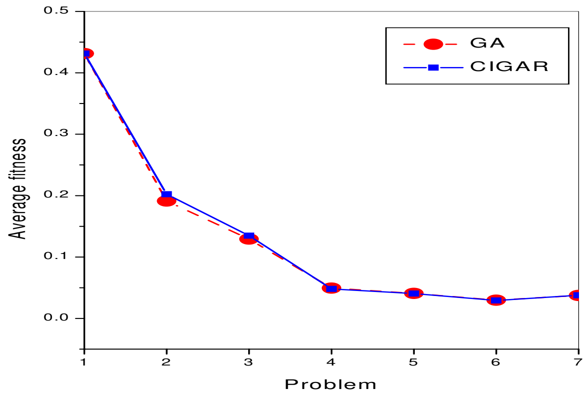

Figure 7: Avg. maximum fitness vs. number of problems attempted

Figure 8 compares the time taken to find these best solutions by the GA versus the time taken by CIGAR. The x-axis is again the number of spectra analyzed while the y-axis plots the average (over ten runs) time taken to find the best solution. We can see that CIGAR improves upon the time taken to find solutions during the first phase (problems 1, 2, and 3). When the temperature peaks resulting in a very different problem, CIGAR's performance (in terms of speed) suffers since injected cases are not relevant to this problem. However, CIGAR does find a quality solution indicating its robustness under adverse conditions. During the second phase performance once again improves, until it overtakes the GA on the last problem.

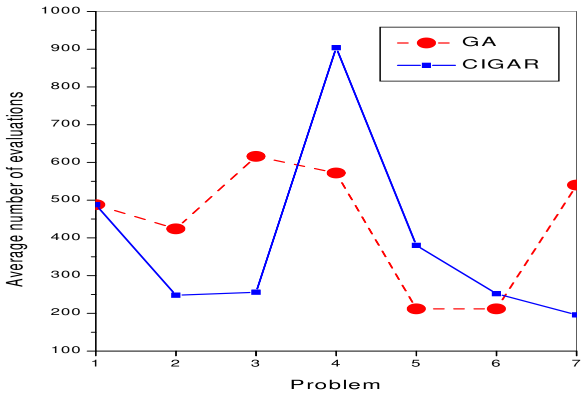

Figure 8: Avg. convergence time vs number of problems attempted

4 CONCLUSIONS AND FUTURE WORK

We studied the possibility of using genetic algorithms for the analysis of experimental X-ray spectra and applied the GA to many experimental spectra with promising results. The GA reliably found good solutions in a relatively small number of evaluations. We emphasize that the plasma temperatures and densities obtained with the GA-driven spectra analysis are consistent with values expected at the collapse of an ICF implosion. We also applied case injected genetic algorithms to seven time-resolved spectra. Our results indicate that when the spectra are similar, CIGAR learns to take less time to find good solutions. When problems are dissimilar, CIGAR is robust enough to recover and find acceptable solutions although it may take longer. Along with parallelization, CIGAR should significantly reduce computing time allowing for the analysis of large numbers of spectra.

We are planning to use our methodology to analyze experimental data using more complete theoretical models. We expect that for optically thick plasma, the temperature and density diagnostics may be ambiguous and this might cause the evaluation function to have a multi-peak structure. This could make the problem less suitable for other less robust algorithms and even more suitable for genetic algorithms and CIGAR. Also, for situations involving many plasma parameters (e.g., parameterization of plasma gradients, background level analysis) the availability of efficient search algorithms for data analysis is critical.

Our preliminary work in this area with more complex but more accurate physical models (optically thick non-uniform plasma) shows promise. Both the GA and CIGAR perform well if we increase the population size and number of generations to 50 to cater to the larger search space brought on by the extra parameters in the more complex physical model. Unfortunately, detailed analysis of performance and speed does not seem to be feasible at this point since fitting one spectrum takes several days.

Acknowledgments

This material is based upon work supported by the National Science Foundation under Grant No. 9624130, and in part by NSF Grant No. OSR-9353227.

References

Eshelman, L. J. (1991). The CHC adaptive search algorithm: How to have safe search when engaging in nontraditional genetic recombination. In Rawlins, G. J. E., editor, *Foundations of Genetic Algorithms-1*, pages 265–283. Morgan Kauffman.

Goldberg, D. E. (1989). *Genetic Algorithms in Search, Optimization, and Machine Learning*. Addison-Wesley.

Golovkin, I. E. and Mancini, R. C. (1999). High order satellites and plasma gradients effects on the Ar He-beta line opacity and intensity distribution. *Journal Of Quantitative Spectroscopy And Radiative Transfer*, In press.

Griem, H. R. (1992). Plasma spectroscopy in inertial confinement fusion and soft x-ray laser research. *Plasma Physics*, 4:2346 – 2361.

Haynes, D. A. and et. al. (1996). Effects of ion dynamics and opacity on Stark-broadened argon line profiles. *Physical Review E*, 53:1042–1050.

Holland, J. (1975). *Adaptation In Natural and Artificial Systems*. The University of Michigan Press, Ann Arbour.

Louis, S. J. and Johnson, J. (1997). Solving similar problems using genetic algorithms and case-based memory. In *Proceedings of the Seventh International Conference on Genetic Algorithms*, pages 283–290. Morgan Kauffman, San Mateo, CA.

Louis, S. J. and Rawlins, G. J. E. (1991). Designer genetic algorithms: Genetic algorithms in structure design. In *Proceedings of the Fourth International Conference on Genetic Algorithms*, pages 53–60. Morgan Kauffman, San Mateo, CA.