Cluster analysis of the hot subdwarfs in the PG survey.

by Peter Thejll[1], Darryl Charache, Harry L. Shipman
Department of Physics and Astronomy, University of Delaware,
Newark, DE 19716.

Introduction

The Palomar Green survey (Green et al., 1986) of faint blue, high galactic-latitude objects, turned up several interesting new classes of objects, such as gravitational lenses (Weyman et al., 1980), the ultra hot star H1504+65 (Nousek et al., 1986), and the PG 1159-035 variables (McGraw et al., 1979). The PG survey forms a mainstay in the investigations of late stages of stellar evolution. Work (Flemming et al., 1986) has already cast light on the important question of the space density of DA's and continuing work is seeking values for similar population parameters for the subdwarfs: Evolutionary links between the subdwarf stage and white dwarfs will thereby be illuminated.

Astrophysics shares certain traits with botany and zoology - the collection and counting of different types of objects. The reason is straightforward; it is hoped that different types of stars represent different evolutionary stages or that the relative distribution over the classes of objects will give insight into evolutionary rates etc. In biology it has happened that nearly identical specimens from the same species were 'classified' as representing a new species - only to turn out to be the male or the female of the species. The same might happen in the astrophysical beastiary if too fine distinctions are made between objects - notably if the distinctions are based on data of low quality.

If the subdivisions are objectively defined and if the data is of such a quality that the categories are significantly different from each other then one may be able to use the characteristics of these classes to study evolution. The key is to choose objective criteria for subdividing the spectra. No matter how important the chosen criteria for subdivision are, nothing will be learned if the data is of such low quality that the groupings are sensitive to noise in the data. Cluster analysis, carried out properly, offers the means for such a robust subdivision.

Using cluster analysis

Cluster analysis consists of certain arithmetic steps carried out on numerical data. The clusters are found as algorithms look for similarity between data points based on a 'distance' between the objects. The rules for calculating the distance and the rules for deciding on the degree of similarity are chosen from a large set of possible methods. One problem with using cluster analysis is that the results can depend on the method used. The best

[1]This research was supported by the Danish Research Academy, grant no. 880070, the NSF grants AST-8720, AST-8515747 and NASA grant NAG 5-972.

choice is one that gives results that are *similar* to those of other methods. You should not choose a method that gives results *different* from most other methods - the information would be artefacts of the method and not consequences of the data.

There is furthermore the question of how many clusters to subdivide the data into. Most clustering programs will give a dendrogram, or a tree, showing the interrelations of the data points. On a dendrogram (Figure 1) you can see how closely related any two classes are by how far up the tree they are joined by a horizontal branch - the higher, the less related are the two groups. At one end of the tree all objects fall into two groups and at the other end they fall into individual groups with one member in each. Where should one 'cut' the dendrogram? If the data was noise free one could choose the cut anywhere one wished purely from considerations of the ultimate needs of the user; any degree of smoothing or enhancement of details could be furnished. When the data is noisy, however, it becomes meaningless to cut the tree at a level so low that, given typical noise, many of the objects might as well be in other clusters. Usually the user wants as much detail as possible so the cutoff point is defined by the level at which noise does not significantly alter the memberships of the clusters.

The Data

We used values for the equivalent widths of the lines H_γ at 4340 A, HeI at 4387 A, HeI at 4471 A, HeII at 4542 A, HeII at 4686 A and HeI at 4713 A. These values were obtained from spectra taken of the subdwarfs with the 90 inch telescope at Kitt Peak by Dr. Richard Green (Kitt Peak) and Dr. James Liebert (Steward Observatory) during several years. Between the spectra quality varies: The best are few and have very little noise, there are about as many of low quality with barely visible lines while the majority have medium quality and often allow the determination of the weakest lines' equivalent width to within 50% or so. The equivalent widths were obtained partly at Kitt Peak and at Steward Observatory using software that did the calculations from files of the spectra on

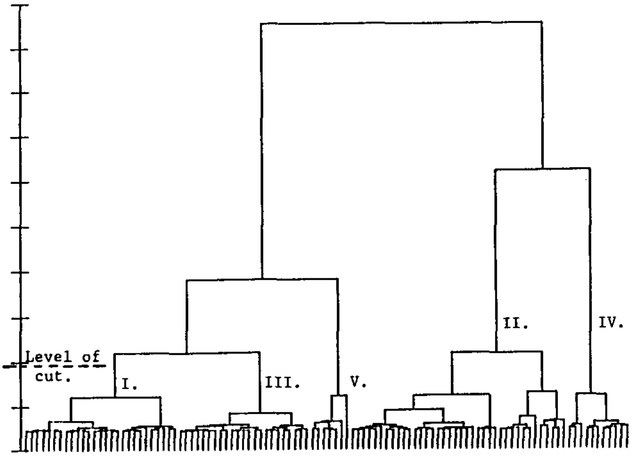

Figure 1: A dendrogram of the average equivalent widths of the Hydrogen, Helium I and Helium II lines in the spectra of 106 hot, helium rich spectra. If the tree is cut at the level shown 5 groups appear. This is the lowest level at which the tree can be cut, given the typical noise in the data. The tree should not be cut lower so that group II breaks into two classes. The dendrogram was produced using Euclidean distances between the data points and clustered by Wards method.

https://doi.org/10.1017/S0252921100099425 Published online by Cambridge University Press

magnetic tapes and partly at University of Delaware, using a hand digitizer and photocopies of plots of the spectra. Although distortions are introduced by this procedure the errors are in no cases as large as the noise inherent to the observational data.

The spectra we have form a subset of the complete survey of subdwarfs in the PG catalog. Some of the spectra were clearly from high gravity objects with line widths of a hundred Angstroms, some were of medium gravity and some were from objects originally classified as white dwarfs in the PG survey before higher quality spectra were obtained. We choose to apply cluster analysis to all Helium rich objects of medium gravity. This is 106 stars out of 121. Liebert and Green have taken pains to obtain spectra of good quality from as many of the 'sdO-like' stars in the PG survey as possible; there should be no selection effects in the subset we have. There are about 220 sdO's in the PG catalog: Applying a correction for the white dwarfs reclassified as sdO's, we see that our subset of 106 sd's is about 50% of all the subdwarfs in the volume of stars surveyed. As has been pointed out (Green et al., 1986) the survey is *complete* for subdwarfs - the observed cone stretches outside the galaxy and has not missed any subdwarfs.

It became clear that we should average all lines from similar ionic species to avoid dominance of line types that appear often - like HeI lines. Therefore we averaged all Hydrogen lines, all HeI lines and all HeII lines and used these averages for the cluster analysis.

The Results

We applied cluster analysis, using the CLUSTAN program (Wishart, 1982), to the above data. Distances were calculated using the Euclidean formula and clustering was done by Wards method. Based on our experience with the methods available in CLUSTAN we feel confident in stating that, given spectra of the available quality, the following 5 groups represent natural divisions of the subdwarf stars in the PG survey.

GROUP	ABUNDANCE	FEATURES
I	(26 %)	**H> HeI and HeII.** H = 5 +/- 1 A, HeI and HeII = 1 +/- 1 A.
II	(36 %)	HeII = 3 +/- 1 A, H = 2 +/- 1 A, HeI = 1 +/- 1 A.
III	(22 %)	H = 3 +/- 1 A, HeII = 1 +/- 1 A, HeI nearly absent.
IV	(10 %)	**HeI >> H and HeII.** HeI = 4 +/- 1 A, H and HeII nearly absent.
V	(7 %)	**H >> HeI > HeII.** H = 9 +/- 1 A, HeI = 2 +/- 1 A, HeII nearly absent.

Here H represents the average of visible Balmer lines, often the β,δ lines and always the γ line. HeI is the average of 4387 and 4471 and sometimes 4026 and 4713. HeII is the average of 4686 and 4542. The given deviations are +/- 1 standard deviation of the mean and do not define upper and lower limits.

We used our scheme to classify sdO's for which NLTE analysis has been published (Hunger et al., 1981) and found preliminary group characteristics for groups **I** and **II**: Group **I** has $T_{eff} \approx$ 38 000K +/- 1500K, log(g) $\approx$ 5.9 +/- 0.5 and composition y $\approx$ 0.1 +/- 0.05. Group **II** has $T_{eff} \approx$ 48 000K +/- 5000K, log(g) $\approx$ 5.6 +/- 0.5 and y $\approx$ 0.4 +/- 0.15.

Discussion

The above groups only coincide with the PG classes to some extent. Notably group **V** is like the sdOA's and group **IV** is like the sdOD's. The stars in these two groups have spectra that all are very similar. The stars in the three

https://doi.org/10.1017/S0252921100099425 Published online by Cambridge University Press

other groups all show evidence of temperature differences. Group **I** is dominated by Hydrogen lines but shows a sequence of additional He lines: HeII only, HeII+HeI and HeI only. Since the H lines are similar between these three subgroups composition must be fairly constant and T_{eff} varies. Groups **II** and **III** show the same pattern.

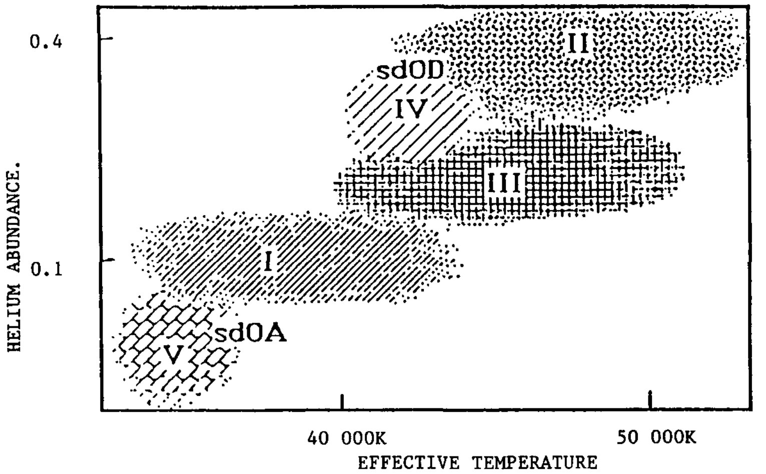

Figure 2: T_{eff} vs. Helium abundance for the five clusters found in the cluster analysis of the subdwarfs. Groups **IV** and **V** are well defined while groups **I**,**II** and **III** show a range of temperatures at roughly constant composition.

References

Green,R., Schmidt, M., and Liebert, J. The Palomar-Green Catalog of Ultraviolet-Excess stellar objects. *Astrophysical Journal Supplement Series*, 1986, *61*, 305-352.

Weyman, R.J., Latham, D., Angel, J.R.P., Green, R.F., Liebert, J.W., Turnshek, D.A., Turnshek, D.E., and Tyson, J.A. The triple QSO PG1115+08: another probable gravitational lens. *Nature*, 1980, *285*, 641.

Nousek, J.A., Shipman, H.L., Holberg, J.B., Liebert, J.W., Pravdo, S., White, N.E., and Giommi, P. H1504+65: An Extraordinary Hot Compact Star Devoid of Hydrogen and Helium. *Astrophysical Journal*, 1986, *309*, 230-240.

McGraw, J.T., Starrfield, S.G., Liebert, J., and Green, R. PG1159-035: A New, Hot, non-DA Pulsating Degenerate. In H. Van Horn and V. Weidemann. (Eds.), *Proceedings of IAU Colloquium No. 53: White Dwarfs and Variable Degenerate Stars.* IAU, 1979.

Flemming, T.A., Liebert, J., Green, R.F. The luminosity function of DA white dwarfs. *Astrophysical Journal*, 1986, *308*, 176-189.

Wishart, D. *CLUSTAN User Manual* (Third ed.). Inter-University/Research Councils Series, Report No 47, 1982.

Hunger, K., Gruschinske, J., Kudritzki, R.P., Simon, K.P. NON-LTE Analysis of Subluminous O-stars. IV Spectral Photometry and NLTE Analysis of 11 Subluminous Stars. *Astronomy and Astrophysics*, 1981, *95*, 244-249.

https://doi.org/10.1017/S0252921100099425 Published online by Cambridge University Press