SPECTRAL CLASSIFICATION, PHOTOMETRY AND STATISTICAL ANALYSIS

A. Heck

Astronomy Division, ESTEC, Villafranca Satellite
Tracking Station, E.S.A., Spain

1. GENERAL CONSIDERATIONS

Sophisticated algorithms of modern statistical analysis are not yet widely applied to astronomical data. For the moment, these methods are mainly used by scientists in fields where the physical basis is less well determined than in astronomy. The algorithms could however be quite useful in new fields of our science where the physics is still being investigated. In more classical applications we have found them quite useful in the study of the relationship between photometric and spectroscopic data.

The photometric catalogue we considered convenient for our purpose was the compilation of uvbyβ measurements by Hauck and Lindemann (1973), interesting because of the great amount of data it contains and because of the homogeneity of the data (Lindemann, 1974). We retained only stars for which all indices (b-y), m_1, c_1, and β had been measured and which received a MK spectral type by Oblak et al. (1976a). In order to deal only with "normal stars" in those first attempts, all peculiar stars (emission-line, magnetic and metallic-line stars) were rejected.

2. LINK WITH A PHYSICAL PARAMETER

First we applied various methods of multivariate statistical analysis to the sample of remaining stars in order to relate a photometric index or a group of photometric indices to a fundamental physical parameter (Heck, 1976). This parameter was chosen to be the effective temperature, through the spectral

https://doi.org/10.1017/S0252921100104166 Published online by Cambridge University Press

type, and although almost no physical hypothesis was used, the results were in full agreement with those obtained by physical considerations for the three groups of spectral types (early, intermediate, late) introduced by Strömgren (1966, 1967).

More precisely the method of selection of variables in discriminant analysis allowed us, with only the notion of categories, to relate numbers of indices to the spectral type or the effective temperature. It permitted also to detect at which point additional indices did not lead to a relevant increase of discrimination. The method of multiple stepwise regression proved to be the most efficient way to quantify the coefficients of linear combinations of the groups of indices retained. The discrimination capacity of the relations obtained was very close to that of the predictors given by Strömgren (1966, 1967).

Such an analysis performed with the effective temperature can surely be repeated with another physical parameter (such as the luminosity) and/or with another type of catalogue. Let us insist on the fact that such statistical methods should not replace a physical analysis, but they can be applied usefully to clear up the problem. A physical analysis then can fit the coefficients in order to take account of the interstellar reddening, to adjust scales,...

3. DETECTION OF DISCREPANCIES

A more spectacular use of the link between spectroscopic and photometric data was made by the cluster analysis methods applied to the same sample of stars (Heck et al., 1977). These methods group stars according to a similarity or dissimilarity measure to be defined between them. The (dis)similarities are only functions of the numerical values of the photometric indices. No physical hypothesis was interfering here.

The groups of stars obtained corresponded to an MK spectral classification with a good discrimination for the spectral types; the discrimination for the luminosity classes was slightly worse. 8% of the stars were misclassified and we guess it is due to a wrong spectral type, to poorly determined photometric indices or to this delicate interface between photometry and spectroscopy. Photometric indices, even well selected, will never give as much information as a spectrum. A not negligible number of those stars appear also in Table 1 of Oblak et al. (1976a) (stars showing large discrepancies between their photometric indices and the means of their spectral group) and in a list of suspected stars by Oblak et al. (1976b). See also Gehren (1977) for HD 218396 for instance.

Here again there is no question of replacing the fine analysis by a specialist who would investigate every measure. But let us appreciate the role such statistical methods could play in defining, for instance, a classification from UV spectral features, an area where little has been done up to now. This work could be carried out independently of any photometry, since the methods are applicable to continuous data, as well as discrete or even binary (presence or absence of a feature) data. We are presently investigating this aspect.

4. PREDICTION OF SPECTRAL TYPE

Considering that the result discussed in the previous section indicated that a spectral class could be defined from the photometry with a 92% chance of being correct, we have attempted (in collaboration with G. Mersch) to develop a prediction algorithm. Its main advantage is the rapidity of the collection of photometric data comparatively to spectroscopic ones.

Due to the particular folded shape of the star catalogue representation in a color-temperature diagram or in a principal-component space, we have first proceeded to a multiple stepwise regression of T_{eff} values on powers of the photometric indices in order to "unfold" the polynomial relationship. Then an isotonic regression performed on the selected variables allowed the determination of one prediction index depending essentially on the rank of the spectral type, and not on the value of the underlying effective temperature.

The procedure works quite well for a luminosity class. An algorithm which is achieved should predict luminosity as well. Therefore we have developed an heuristic rule which selects that luminosity class first. It seems now, however, that a better way is to use an empirical determination based on a principal-component analysis.

The best results are obtained, as expected, for main-sequence stars for which the precision of the prediction is about one spectral subtype. For the other luminosity classes some permutations appear between subgroups. As a first approach we merge those subgroups. In spite of the consequent loss of the precision for those classes, the prediction obtained is an excellent approximation.

REFERENCES

Gehren, T. (1977). Astron. Astrophys. 59, 303.
Hauck, B. and Lindemann, E. (1973). Astron. Astrophys. Suppl. 11, 119.
Heck, A. (1976). Astron. Astrophys. 47, 129.
Heck, A., Albert, A., Defays, D. and Mersch, G. (1977). Astron. Astrophys. 61, 563.
Lindemann, E. (1974). Astron. Astrophys. Suppl. 15, 209.
Oblak, E., Chareton, M. and Considere, S. (1976a). Astron. Astrophys. Suppl. 24, 69.
Oblak, E., Chareton, M. and Considere, S. (1976b). private communication.
Strömgren, B. (1966). Ann. Rev. Astron. Astrophys. 4, 433.
Strömgren, B. (1967). in The Magnetic and Related Stars, R. Cameron, ed., Mono Book Corp., Baltimore, p. 461.

DISCUSSION

Cowley: The methods you have been discussing seem ideally suited to the study of spectroscopic peculiarities that one can find on the upper main sequence, where the number of parameters necessary to describe the spectra is obviously large – 25 or more – if one considers high dispersion spectra.

Heck: Of course, one of the advantages of these methods is that you can use as many variables as you wish, since you are working generally in a N-dimensional space. However the more variables you are considering, the more stars your sample must contain to render your results significant from a statistical point of view. Maybe I should add that, according to past experience with, for instance, medical or psychological data, another advantage of these algorithms consists in often pointing out aspects of the data you would not notice by a classical approach.