УДК 524.3—355—17

М. Ш. КАРАПЕТЯН

О ВОЗМОЖНОСТИ СТАТИСТИЧЕСКОЙ СПЕКТРАЛЬНОЙ КЛАССИФИКАЦИИ ЗВЕЗД

Рассмотрена возможность классификации звезд методами многомерного статистического анализа (МСА). Посредством выделения двух главных факторов для заданных стандартных звезд определены спектральные классы, которые с вероятностью, больше 90%, совпали с их известными спектральными классами. Создан пакет программ на ЭВМ СМ-4 для осуществления статистической классификации звезд.

В настоящей работе рассматривается возможность статистической спектральной классификации звезд. Методы многомерного статистического анализа (МСА), которые применяются в работе, до настоящего времени в основном использовались лишь в социальных науках. В астрономии МСА впервые стали применяться в работах [1—3]. В приложениях А и Б приводится краткое описание этих методов, а их более полное описание можно найти в работах [4—6]. Для спектральной классификации используются так называемые спектральные индексы, представляющие собой индикаторы внутренних (температура, поверхностное притяжение, химический состав, вращение и т. д.) и внешних (покраснение, наблюдаемый шум и т. д.) параметров звезд. В МСА можно использовать любые параметры (интенсивность спектральных линий, градиенты излучения в разных частях спектра, излучение в непрерывном спектре и т. д.) звезд. В данной работе применяются два вида данных: спектральные индексы, вычисленные из сканных спектров [7] и абсолютное распределение энергии в спектре [8].

1. В работе [7] для 41 стандартной звезды карлика спектральных классов от поздних O до ранних G в спектральном диапазоне λ 3500—4400 Å приводятся следующие спектральные индексы: H_γ, G-полосы, CaI (4227), CaII (3934), H 3889, БС-бальмеровский скачок и соответствующие индексы континуума (C_1, C_2, C_3, C_4, C_5, C_6). К этим данным нами был применен факторный анализ (приложение А) и в результате выделены два главных фактора—F1, F2 (табл. 1). Как видно из этой таблицы, F1 объясняет 55%, а F2—32% общей дисперсии всех данных. При этом оказывается, что фактор F1 обусловливается спектральными индексами G-полосы, CaI, CaII, C 4, C 5, C 6 и зависит от спектрального класса линейно (рис. 1, *а*). Фактор же F2 обусловливается индексами H_γ, H 3889, БС, C1, C2 и также коррелирует (параболическая зависимость) со спектральным классом (рис. 1, *б*).

После выделения главных факторов (F1, F2) можно провести спектральную классификацию звезд, используя дискриминантный анализ (см. приложение Б). Для этой цели 34 стандартные звезды из 41 были разделены на семь классов, остальные же семь звезд из разных классов заново были классифицированы. В результате полученные классы у всех семи звезд с вероятностью больше 90% совпали со своими фактическими спектральными классами [8].

Таблица 1

Значения факторных нагрузок и общих дисперсий в случае с 12 индексами

№	Индекс	Фактор F1	Фактор F2
1	H_γ	—0.336	0.910
2	G—полоса	0.895	—0.375
3	CaI, λ 4227	0.901	—0.213
4	CaII, λ 3934	0.960	—0.081
5	H_ε, λ 3899	—0.322	0.918
6	B1—бальмеровский скачок	0.042	0.953
7	C1	0.442	—0.867
8	C2	—0.200	—0.856
9	C3	—0.311	—0.438
10	C4	—0.961	—0.017
11	C5	—0.990	—0.038
12	C6	—0.976	—0.057
Дисперсия, %		55	32

2. В книге [8] приводятся данные о распределении энергии в спектрах звезд разных спектральных классов и классов светимости в диапазоне длин волн λ 3200—7600 А. Из этих спектральных данных нами определялись спектральные индексы, близкие к фотометрическим индексам Вильнюсской многоцветной фотометрической системы [9] (U, P, X, Y, Z, V, S) и систем Гильденкерна [10] (g, n, f, u). В табл. 2 и 3 приводятся спектральные индексы в двух интервалах длин волн: а) λ 3400—6600 А и б) λ 3700—5500 А (табл. 2).

Для 150 стандартных звезд III и V классов светимости спектральных классов от M до O нами были вычислены 16 спектральных индексов для интервала «а» и 10 для интервала «б». Применение факторного анализа (приложение А) позволило выделить (для интер-

Таблица 2

Значения факторных нагрузок и общих дисперсий в случае с 16 индексами

№	Индекс	Фактор F1	Фактор F2	Фактор F3
1	$E_{3450}-E_{4050}$	—0.348	0.898	—0.013
2	$E_{3450}-E_{3750}$	—0.034	0.878	0.044
3	$E_{3750}-E_{4050}$	0.540	0.796	—0.072
4	$E_{4050}-E_{4650}$	0.938	0 290	—0.038
5	$E_{4650}-E_{5150}$	0.942	0.082	—0.052
6	$E_{4050}-E_{5150}$	0.961	0.221	—0.043
7	$E_{5150}-E_{6550}$	0.960	—0.034	—0.069
8	$E_{4275}-E_{4400}$	0.932	0.166	—0.049
9	$E_{4050}-E_{3900}$	—0.444	—0.758	0.037
10	$E_{4050}-E_{4975}$	0.952	0.241	—0.042
11	$E_{5150}-E_{5450}$	0.931	0.060	—0.100
12	$E_{5440}-E_{6550}$	0.934	—0.082	—0.050
13	E_{3450}	0.702	0.703	0.029
14	E_{3900}	0.927	0.343	0.029
15	E_{4500}	—0.132	—0.011	0.988
16	E_{6550}	—0.911	0.075	0.260
Дисперсия, %		66	19	6

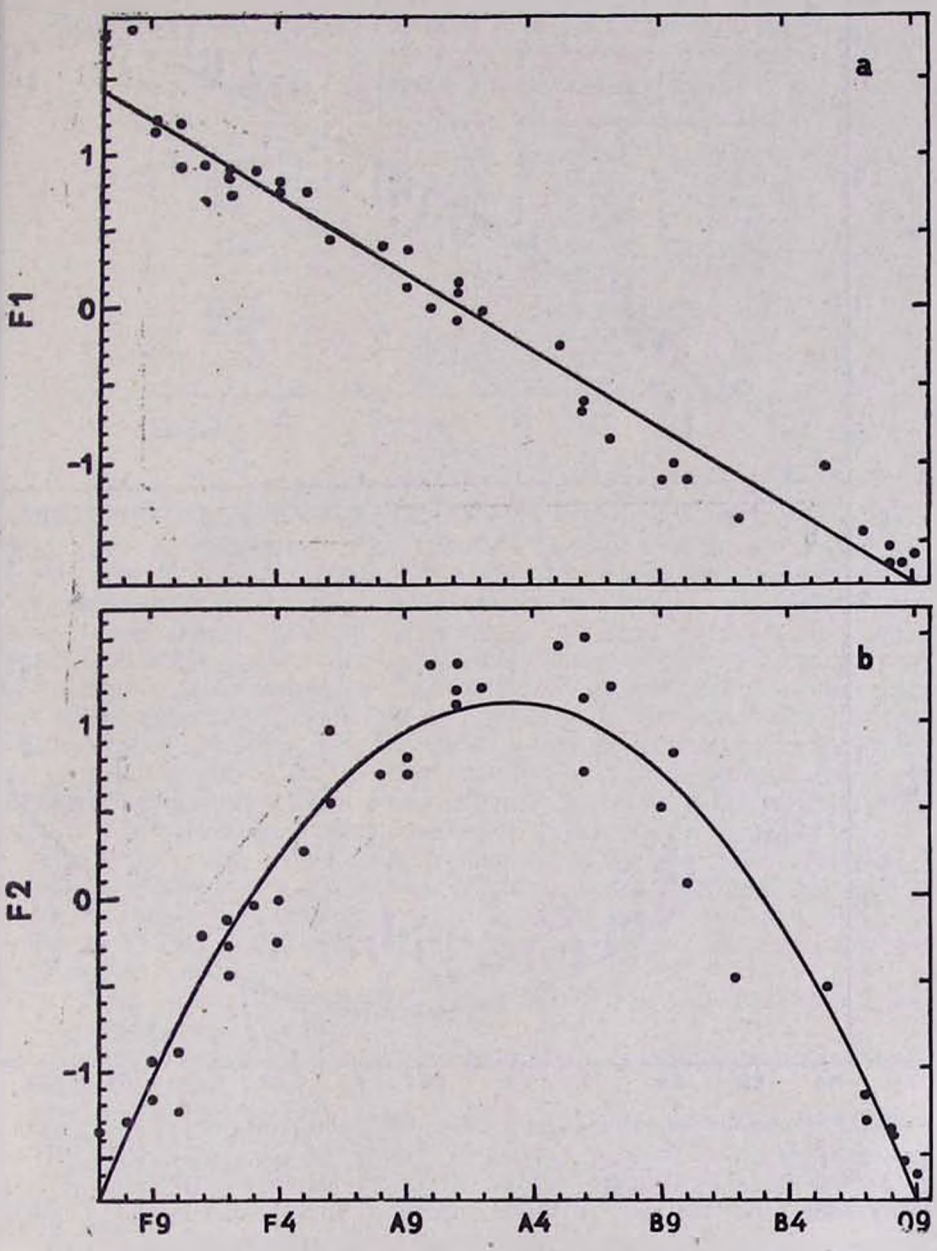

Рис. 1,*a*—зависимость спектрального класса звезд от фактора F1. Сплошная линия приближает положения звезд полиномом 1-й степени; *b*—зависимость спектрального класса звезд от фактора F2. Сплошная линия приближает положения звезд полиномом 2-й степени

вала «а») три главных фактора F1, F2, F3 (табл. 2). Как видно из табл. 2, фактор F1 объясняет 66%, F2—19%, F3—6% общей дисперсии всех данных. Фактор F3 не коррелирует ни с каким из звездных параметров и ввиду его малой доли в общей дисперсии его вкладом

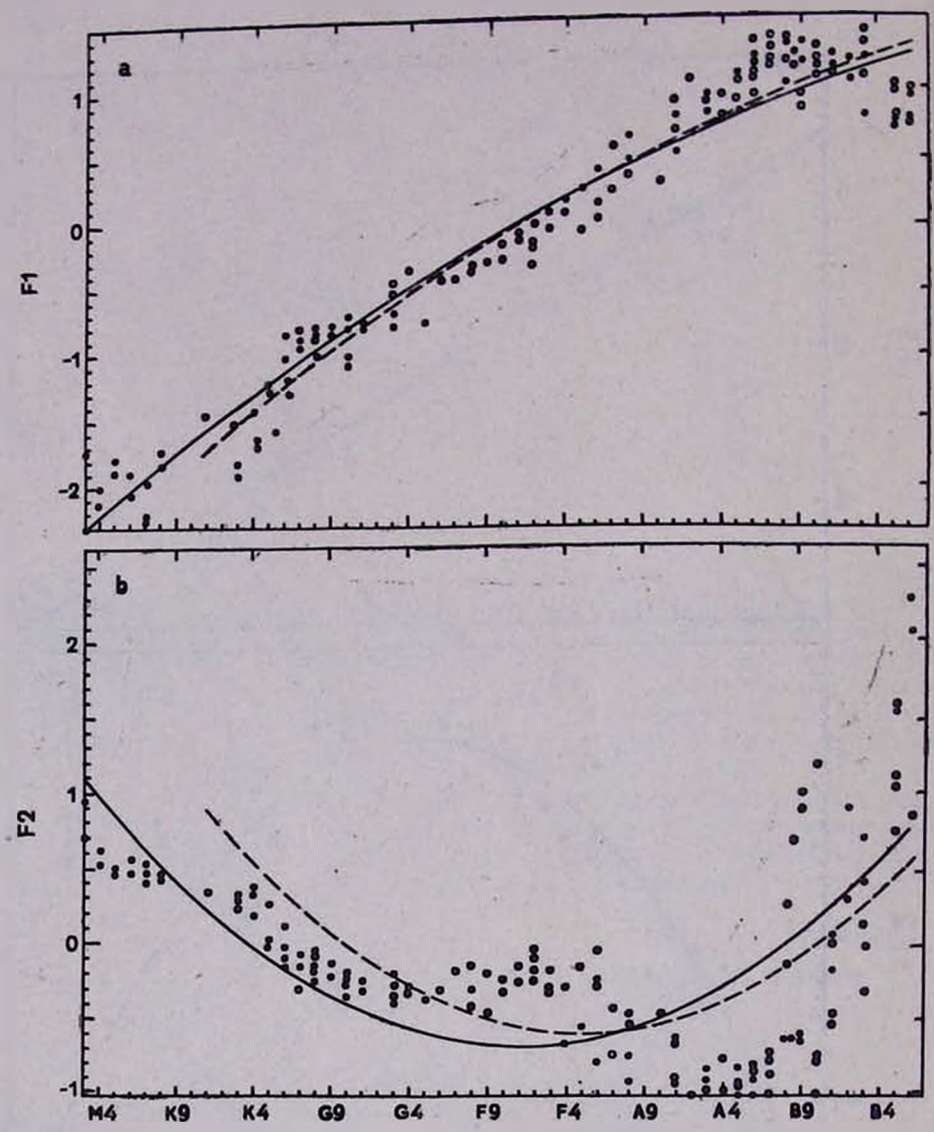

Рис. 2,*a*—зависимость спектрального класса звезд от фактора F1; *b*—зависимость спектрального класса звезд от фактора F2. Знаком (°) обозначены звезды V класса светимости, а знаком (●)—звезды III класса светимости. Прерывистая линия приближает точки для звезд V класса светимости полиномом 3 степени. Сплошная линия—то же для звезд III класса светимости

можно пренебречь. При этом ясно, что факторы F1 и F2 коррелируют довольно хорошо со спектральными классами (рис. 2, *a* и 2.*b*).

Аналогичный анализ для интервала «б» также позволяет выделить три главных фактора F1, F2, F3 (табл. 3). Как видно из табл. 3, F1 объясняет 84%, F2—10%, F3—3% общей дисперсии всех данных. Здесь, как и в случае интервала «а», по причине малого вклада

Таблица 3

Значения факторных нагрузок и общих дисперсий в случае с 10 индексами

№	Индекс	Фактор F1	Фактор F2	Фактор F3
1	$E_{3750}-E_{4050}$	0·330	−0·880	0.255
2	$E_{4050}-E_{4450}$	0.853	−0.464	0.207
3	$E_{4050}-E_{5150}$	0.947	−0.236	−0·030
4	$E_{4050}-E_{5150}$	0.905	−0.398	0.133
5	$E_{4375}-E_{4400}$	0.890	−0.370	0.094
6	$E_{4050}-E_{3900}$	−0.290	0·937	0·054
7	$E_{4050}-E_{4375}$	0·681	−0.418	0.189
8	$E_{5150}-E_{5450}$	0.812	−0.204	0.531
9	E_{3900}	0.831	−0.521	0.172
10	E_{4500}	0.915	−0.258	0.260
Дисперсия, %		84	10	3

фактором F3 можно пренебречь. Из рис. 3, *а* и 3, *б* видно, что факторы F1 и F2 хорошо коррелируют со спектральными классами.

Из рис. 2, *б*, 3, *б* видно, что помимо корреляции факторов со спектральными классами, обнаруживается также разделение (не очень явное) звезд по классам светимостей. Для более четкого разделения, вероятно, надо использовать большее количество звезд разных классов светимостей, а возможно, и другие наборы индексов. Для классификации звезд (как и в разделе 1) мы использовали дискриминантный анализ. Заново были классифицированы две группы (из интервалов «а» и «б»), содержащие по 12 стандартных звезд. В результате найденные нами спектральные классы с вероятностью 90% совпали с известными спектральными классами этих звезд.

Статистический метод классификации имеет ряд преимуществ:

а) можно использовать возможно большее число спектральных индексов, которые характеризуют звезды, найти среди них наиболее информативные и тем самым повысить точность и качество классификации;

б) для хорошей классификации звезд необязательно получать спектры с большой дисперсией;

в) статистический метод дает возможность классифицировать более слабые звезды;

г) применение ЭВМ дает возможность в короткое время классифицировать большое число звезд.

Следовательно, мы видим, что статистическими методами можно достичь более точной, объективной, однородной и быстрой классификации звезд.

Представленный выше метод предполагается использовать для классификации спектров, полученных с объективной призмой на метровом телескопе системы Шмидта Бюраканской астрофизической обсерватории.

Приложение А—Факторный анализ

В этой части дается краткое описание процедуры, используемой в МСА для нахождения линейно независимых факторов, содержащихся в исследуемом массиве.

Допустим, что Z-массив наблюдаемых величин, в котором $z_{ij} \in Z$, а i-ый индекс нормализован к нулевой средней и единичной диспер-

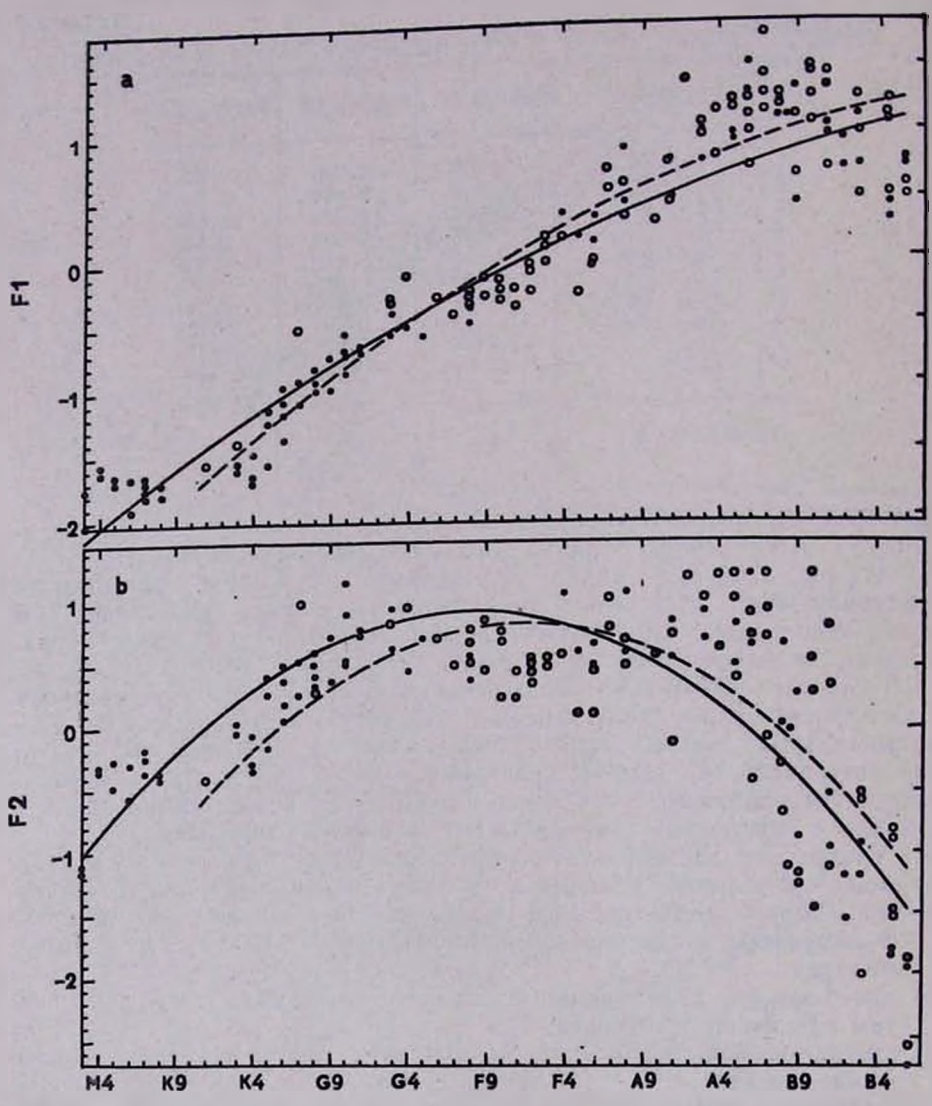

Рис. 3,*a*—зависимость спектрального класса звезд от фактора F1; *b* —зависимость спектрального класса звезд от фактора F2. Обозначения соответствуют обозначениям к рис. 2

сии для j-ой звезды (i=1, 2, ..., m; j=1, 2, ..., n). Матрица Z представляется в виде линейной комбинации r гипотетических факторов.

$$z_{ij}=a_{i1}\ p_{1j}+a_{i2}\ p_{2j}+\ldots+a_{ir}\ p_{rj} \qquad (1)$$

или в матричной форме Z=AP, где A=(a_{ic})—матрица порядка m×r —факторное отображение массива Z, P=(p_{ij})—матрица порядка r×n—значения факторов всех индивидумов.

Матрицы A и P неизвестны, и для их нахождения надо ввести дополнительные условия.

Очевидно, что корреляционная матрица равна

$$R=\frac{1}{n-1}\,Z\,Z'. \tag{2}$$

Подставляя (1) в (2), получим

$$R=\frac{1}{n-1}\,APP'A'=A\,\frac{1}{n-1}\,PP'A'.$$

По аналогии с (2) можно утверждать, что $C=(c_{ef})=\frac{1}{n-1}PP'$ является корреляционной матрицей, отражающей связи между факторами и, следовательно,

$$R=ACA'. \tag{3}$$

Если наложить на (3) условие некоррелированности факторов, т. е. принять $C=1$, то в результате получим

$$R=AA'. \tag{4}$$

Система (4) имеет однозначное решение при вводе следующих дополнительных условий: сумма квадратов нагрузок первого фактора должна составлять максимум от полной дисперсии; сумма квадратов нагрузок второго фактора должна составлять максимум оставшейся дисперсии и т. д., т. е. эти условия максимизируют функцию

$$S_1=\sum_{i=1}^{m} a_{i1}^2=\max. \tag{5}$$

При $m(m-1)/2$ независимых друг от друга условиях, имеем

$$r_{ik}=a_{i1}\cdot a_{k1} \qquad (i,\ k=1,2,\ldots,\ m,\ i<k).$$

Для максимизации функции используется метод множителей Лагранжа. В результате приходим к системе m однородных уравнений с m неизвестными a_{i1} :

$$\begin{aligned}
&(1-\lambda)\ a_{11}+r_{12}\ a_{21}+\ldots+r_{1m}\ a_{m1}=0\\
&r_{21}\ a_{11}+(1-\lambda)\ a_{21}+\ldots+r_{2m}\ a_{m1}=0\\
&\cdot\qquad\cdot\qquad\cdot\qquad\cdot\\
&\cdot\qquad\cdot\qquad\cdot\qquad\cdot\\
&\cdot\qquad\cdot\qquad\cdot\qquad\cdot\\
&r_{m1}\ a_{11}+r_{m2}\ a_{21}+\ldots+(1-\lambda)a_{m1}=0.
\end{aligned} \tag{6}$$

Необходимым и достаточным условием существования нетривиального решения системы (6) является равенство нулю детерминанта матрицы коэффициентов этих уравнений

$$\begin{vmatrix}
(1-\lambda) & r_{12} & \cdots & r_{1m}\\
r_{21} & (1-\lambda) & \cdots & r_{2m}\\
\cdot & \cdot & & \cdot\\
\cdot & \cdot & & \cdot\\
\cdot & \cdot & & \cdot\\
r_{m1} & r_{m2} & \cdots & (1-\lambda)
\end{vmatrix}=0 \tag{7}$$

Если в (6) подставить найденное из (7) значение λ_1, то получим вектор решения $(a_{11},\ldots,a_{m1})$, который удовлетворяет указанным дополнительным условиям и имеет максимум при $\sum_{i=1}^{m} a_{i1}^2$. То же самое

имеет место для λ_2 и т. д. Решение системы (6) сводится к так называемой проблеме нахождения собственных значений действительной симметрической матрицы. В общем виде оно записывается как

$$Rx_e=\lambda_e x_e \quad \text{или} \quad (R-\lambda_e I)x_e=0,$$

где λ_e—собственные значения, соответствующие собственным векторам α_e матрицы R. Путем нормировки получаем искомые значения a_e матрицы A посредством компонент собственных векторов матрицы R:

$$a_{ie}=x_{ie}\sqrt{\lambda_e}\Big/\sqrt{x_{1e}^2+x_{2e}^2+\ldots+x_{me}^2}. \tag{8}$$

Приложение Б—Дискриминантный анализ

Задача классификации состоит в отнесении некоторого объекта ω к одной из k популяций $\omega_1, \omega_2, \ldots, \omega_k$ на основе измерения p параметров $\vec{X}=(X_1, X_2, \ldots, X_p)'$.

Пусть $f_i(\vec{x})$ означает плотность распределения x в ω_i и q_i - априорную вероятность того, что вектор $\vec{x}$ принадлежит популяции ω_i, $i=1, \ldots, k$. Обозначим стоимость отнесения наблюдения из W_j к W_i—через $C(i/j)$, а вероятность отнесения наблюдения из ω_j к ω_i—через $P_r(i/j)$, $i, j=1, \ldots, k$, $i\neq j$. Обобщенная байесовская процедура классификации относит вектор $\vec{X}$ к W_i, если величина

$$-\sum_{\substack{i=1\\ j\neq i}}^{k} q_j f_j(\vec{X})\, C(i/j) \tag{1Б}$$

является максимальной. Величина (1Б) называется значением дискриминантной функции для i-ой популяции. Байесовская процедура минимизирует ожидаемую стоимость ошибочной классификации—

$$\sum_{j=1}^{k} q_j\{\sum_{\substack{i=1\\ j\neq i}}^{k} C(i/j)\ Pr(i/j)\}\ . \tag{2Б}$$

Если полагать все C(i|j) равными, то процедура Байеса отнесет $\vec{X}$ к W_i, если величина

$$q_i f_i(\vec{X}) \tag{3Б}$$

имеет максимальное значение. Таким образом, минимизируется ожидаемая вероятность ошибочной классификации—

$$\sum_{j=1}^{k} q_j \Big\{\sum_{\substack{i=1\\ j\neq i}}^{k} Pr\ (i/j)\Big\}, \tag{4Б}$$

что эквивалентно отнесению $\vec{X}$ к W_i если апостериорная вероятность

$$Pr\ (W_i/\vec{X})=q_i\ f_i(\vec{X})\ /\ \sum_{j=1}^{k} q_j f_j(\vec{X}) \tag{5Б}$$

достигает максимума.

Когда популяция W_i имеет нормальное распределение $N(X_i, S)$ с функцией плотности $f_i(\vec{X}_i)$, $i=1, \ldots, k$, то подставляя $f_i(\vec{X}_i)$,

ив выражение (3Б), логарифмируя и исключая общие множители, получаем линейную дискриминантную функцию для i-ой популяции

$$d_i = a_{i1}\ x_1 + \ldots + a_{ip}\ x_p + C_i + \ln q_i, \qquad (6Б)$$

где

$$\begin{bmatrix} a_{i1} \\ \vdots \\ a_{ip} \end{bmatrix} = S^{-1}\bar{x}_i, \qquad C_i = -\frac{1}{2}\bar{x}_i'\ S^{-1}\ \bar{x}_i,$$

а $\bar{x}_i$ — вектор средних и S — ковариационная матрица оцениваются из выборок.

Вектор $\vec{X}$ классифицируется как принадлежащий популяции W_i если величина d_i имеет наибольшее значение. При этом оценка апостериорной вероятности имеет вид

$$Pr(W_i/\vec{X}) = e^{d_i} / \sum_{j=1}^{k} e^{d_j}. \qquad (7Б)$$

26 июня 1986 г.

Մ. Շ. ԿԱՐԱՊԵՏՅԱՆ

ԱՍՏՂԵՐԻ ՎԻՃԱԿԱԳՐԱԿԱՆ ՍՊԵԿՏՐԱԼ ԴԱՍԱԿԱՐԳՄԱՆ ՀՆԱՐԱՎՈՐՈՒԹՅԱՆ ՄԱՍԻՆ

Վիճակագրական բազմաչափ վերլուծության եղանակների օգնությամբ ցույց է տրվում աստղերի սպեկտրալ դասակարգման հնարավորությունը։ Ալիքային երկարության տարբեր միջակայքերում հաշված սպեկտրալ գործակիցները օգտագործվում են M—O դասի աստղերի սպեկտրալ դասակարգման համար։ Գործոնային վերլուծության (հավելված A) միջոցով առանձնացվել են երկու այն գլխավոր գործոնները, որոնցով անմիջականորեն որոշվում է աստղի սպեկտրալ դասը։ Տարբերիչ վերլուծության միջոցով (հավելված Б) դասակարգվել են մի շարք ստանդարտ աստղեր, որոնց համար ստացված գնահատականները 90%-ից մեծ հավանականությամբ համընկնում են իրենց իրական սպեկտրալ դասերի հետ։ Աստղերի վիճակագրական դասակարգման համար ստեղծվել է ծրագրերի համակարգ CM—4 ԷՀՄ-ի վրա։

M. SH. KARAPETIAN

ON THE POSSIBILITY OF STATISTICAL SPECTRAL CLASSIFICATION OF STARS

The possibility of classification of the stars by methods of multivariate analysis (MVA) is shown. For the classification of the stars with

M—O spectral types the spectral indexes in a different wavelength re gions are used. The use of factor analyses (appendix A) gives two majo factors, which are correlated with the spectral types of the stars. Stan dard stars were classified by discriminant analysis (appendix B) as well The estimated spectral types of all stars with probability greater than 90% coincide with their known spectral types. The package of the prog rammes for the statistical classification of the stars on the SM—4 compu ter is written.

ЛИТЕРАТУРА

1. *T. J. Deeming*, Mon. Not. Roy. Astron. Soc. 120, 267. 1964.
2. *C. Christian, K. A. Janes*, Publ. Astron. Soc. Pacific 89, 415. 1977.
3. *S. M. Faber*, Astrophys J., 179, 731. 1973.
4. *К. Иберла*, Факторный анализ, «Статистика», 1980.
5. *А. Афифи, С. Эйзен*, Статистический анализ. Подход с использованием ЭВМ Мир, 1982.
6. *M. C. Kendall*, A Course in Multivariate Analysis London: Griffin. 1957.
7. *W. I. Beavers D. B. Cook*, Astrophys J. Suppl. Ser., 44, 489. 1980.
8. Спектрофотометрия ярких звезд: Справочник /Под редакцией И. Н. Глушневой/. М., Наука, 1982.
9. *В. Страйжис*, Многоцветная фотометрия звезд. Вильнюс, Изд-во «Мокслас», 1977.
10. *K. Gyldenkerne*, Publ. Kobenhavn obs., № 178, 1964.