

В.В. Гольяпин

Q-ТЕХНИКА И ИТЕРАЦИОННЫЙ МЕТОД ГЛАВНЫХ ФАКТОРОВ В ДИАГНОСТИКЕ СЕРДЕЧНО-СОСУДИСТОЙ СИСТЕМЫ

Необходимость разбиения совокупностей объектов на однородные группы возникает в медицине, психологии, антропологии, лингвистике, археологии и т.д. При классификации объектов или индивидуумов практически отсутствует априорная информация о распределении измерений внутри классов. В то же время стандартные алгоритмы и статистические критерии, позволяющие производить оптимальное разбиение на классы, ориентируются на определенную априорную информацию.

Техника **Q** не требует априорной информации, она исходит из матрицы коэффициентов корреляций между индивидуумами [1]. Определяется простая факторная структура, и с помощью **Q**-техники выделяются группы, внутри которых индивидуумы тесно связаны между собой. Таким образом, факторы, которые выделяются с помощью **Q**-техники, определяют собой классы индивидуумов.

Алгоритм работы **Q**-техники:

1. Формируется матрица исходных данных **Y** типа таблицы объект- свойство размерности $m \times n$, где m — количество параметров, n — количество индивидуумов.
2. Путем элементарного преобразования получаем из матрицы **Y** матрицу стандартизованных данных **Z** размерности $m \times n$.
3. Вычисляем корреляционную **R** матрицу размерности $n \times n$.

$$\mathbf{R} = \frac{1}{n-1} \mathbf{Z}\mathbf{Z}^T. \quad (1)$$

4. Прямым итерационным методом вращения находим матрицу собственных значений Λ и матрицу собственных векторов **U** корреляционной матрицы **R** [2].

5. Методом главных компонент (МГК) с использованием полученных выше матрицы собственных значений и матрицы собственных векторов получаем матрицу факторного отображения \mathbf{A} размерности $m \times r$, r — количество факторов (в \mathbf{Q} -технике классов):

$$\mathbf{A} = \mathbf{U} \cdot \Lambda^{1/2}. \quad (2)$$

6. На этом шаге подвергнем матрицу \mathbf{A} вращению, используя варимакс-критерий:

$$s^2 = \frac{1}{m} \sum_{p=1}^r \sum_{j=1}^m b_{jp}^4 - \frac{1}{m^2} \sum_{p=1}^r \left(\sum_{j=1}^m b_{jp}^2 \right)^2 \rightarrow \max, \quad (3)$$

где

$$\begin{aligned} b_{jp} &= a_{jp} \cos \varphi + a_{jq} \sin \varphi; \\ b_{jq} &= -a_{jp} \sin \varphi + a_{jq} \cos \varphi. \end{aligned}$$

Под φ понимается угол вращения φ_{pq} в плоскости векторов p и q . При ортогональном преобразовании в плоскости весовых нагрузок исходные элементы матрицы \mathbf{A} переходят в новые элементы матрицы \mathbf{B} [3]. В результате вращения получаем простую факторную структуру.

7. Оставляем те факторы, для которых $\lambda_i \geq 1$.
8. Проводим классификацию по следующему принципу: индивидуум принадлежит к тому фактору (классу), значения нагрузок которого больше. Примером может служить следующая оптимальная матрица весовых нагрузок (Таблица 1.).

Таблица. 1. Пример оптимальной матрицы весовых нагрузок, полученной в результате работы \mathbf{Q} -техники.

ФИО индивидуума	Фактор 1 (Класс 1)	Фактор 2 (Класс 2)
Петров В. В.	0,044613	0,997282
Кузнецов А. К.	0,999520	0,028029
Логинов Л. К.	0,007446	0,998297
Корнеев Г. А.	0,999551	0,028257
Васильев П. Р.	0,020088	0,999692
Гуреев А. С.	0,999494	0,003940
Миронов Е. Ш.	0,003254	0,999240

Данным алгоритмом \mathbf{Q} -техники производился анализ таблицы «объект-свойство» с объемом данных 475 индивидуумов. Измерялись антропометрические параметры — рост, вес индивидуумов, параметры сердечно-сосудистой системы — амплитудные и временные характеристики зубцов P, R, T, интервалы PQ, QT и длительность QRS.

В результате получена \mathbf{Q} -структура, которая позволяет разделить генеральную совокупность на три независимых класса — нормостеники, гиперстеники, астеники, т.е. положение сердца в грудной клетке является основополагающим критерием разделения индивидуумов на классы.

Каждый из классов подвергнут обработке итерационным методом главных факторов (МГФ) с поиском оптимальной факторной структуры на группах условно здоровых индивидуумов. Задача заключалась в нахождении простой факторной структуры для каждого класса и построения факторных диаграмм как условно здоровых индивидуумов, так и индивидуумов с патологией сердечно-сосудистой системы.

Отправной точкой МГФ является \mathbf{R}_h — редуцированная корреляционная матрица, на главной диагонали которой стоят не единицы, а общности. В отличие от МГК в итерационном МГФ факторные нагрузки вычисляются последовательно друг за другом и число факторов ограничивается требованиями соответствующей проблемы. Итеративный процесс начинается с выбора вектора $u^{(1)}$, элементы которого являются первыми приближениями значений элементов собственных векторов. Вектор $u^{(1)}$ перемножается с матрицей \mathbf{R} по формуле (4). Разделив элементы результирующего вектора $\tilde{u}^{(1)}$ на наибольший по величине элемент этого же вектора (5), получаем новый вектор $u^{(2)}$, с которым опять повторяется процедура (4). Верхние индексы в скобках означают здесь шаг итерации.

$$\tilde{u}^{(1)} = \mathbf{R}_h \cdot u^{(1)}, \quad (4)$$

$$u^{(2)} = \tilde{u}^{(1)} / \max(\tilde{u}_i^{(1)}). \quad (5)$$

.....

$$u^{(k+1)} = \tilde{u}^{(k)} / \max(\tilde{u}_i^{(k)}). \quad (6)$$

Процесс повторяется до тех пор, пока не добиваются сходимости к первому собственному значению $\lambda_1 = \max(\tilde{u}_i^{(k)})$ и соответствующему первому собственному вектору $u^{(k)}$ матрицы \mathbf{R}_h . Формула (6) является общей для k -шагов. Итеративный процесс заканчивается, когда $u^{(k)}$ и $u^{(k-1)}$ с достаточной точностью совпадают друг с другом. В качестве элементов вектора $u^{(1)}$ используются величины, пропорциональные суммам элементов строк матрицы \mathbf{R}_h . Исходя из полученных значений элементов собственного вектора и собственного значения по формуле (7) вычисляются нагрузки первого фактора $a_1 = (a_{11}, \dots, a_{m1})$, которые затем служат для определения $\mathbf{R}_+ = a_1 \cdot a_1'$.

$$a_{ij} = u_{ij} \cdot \sqrt{\lambda_j} / \sqrt{u_{1j} + u_{2j} + \dots + u_{mj}}. \quad (7)$$

Матрица \mathbf{R}_+ является матрицей коэффициентов корреляции, вычисленных с учетом только первого фактора. Остаточная матрица в общем виде определяется так: $\mathbf{R}_1 = \mathbf{R}_h - \mathbf{R}_+$.

Если принимают решения вычислять вторые факторные нагрузки, то повторяется аналогичная процедура с \mathbf{R}_1 до получения второго собственного значения и второго собственного вектора. Процесс выделения факторов можно продолжить и далее.

Получение редуцированной матрицы осуществляется следующим образом. Принимается решение о выделении определенного числа факторов. Затем выбирают предварительные оценки общностей, в качестве которых могут быть использованы коэффициенты множественной корреляции (КМК). Значения КМК для каждой переменной вычисляются по формуле:

$$r_{ii}^h = 1 - \frac{1}{r_{ii}}, \quad (8)$$

где r_{ii}^{ii} - диагональный элемент матрицы \mathbf{R}^{-1} , а r_{ii}^h - диагональный элемент матрицы \mathbf{R}_h . С помощью метода главных факторов из матрицы \mathbf{R}_h выделяется g столбцов матрицы весовых нагрузок факторов. По полученной матрице вычисляются новые оценки общностей. Вновь выполняется процедура выделения g столбцов матрицы \mathbf{A} с помощью МГФ по матрице \mathbf{R}_h с новыми диагональными элементами. Процесс повторяется до тех пор, пока вычисленные диагональные элементы не перестанут изменяться от итерации к итерации. К полученной на последнем шаге матрице \mathbf{A} применяют варимакс-критерий и получают оптимальную матрицу весовых нагрузок факторов (Таблица 2).

Таблица. 2. Оптимальная матрица весовых нагрузок факторов, характеризующих электрокардиодинамику.

Параметр	Фактор 1	Фактор 2	Фактор 3	Фактор 4	Фактор 5
P, мм	-0,119437	0,045152	0,855707	-0,149865	-0,348385
P, сек.	-0,099755	0,085888	0,496010	-0,804163	0,135162
PQ, сек.	-0,747951	0,224936	0,068552	0,358900	-0,336386
Q, мм	0,714988	-0,024886	0,615811	0,112613	0,019329
R, мм	0,348768	0,090260	0,782967	0,182725	0,040371
S, мм	-0,093307	-0,149977	-0,215281	-0,852451	-0,012823
QRS, сек.	0,036936	-0,019913	0,138620	0,029331	-0,946879
ST, сек.	-0,505610	0,506232	0,112452	0,522348	0,280279
T, мм	0,781202	0,030311	0,119850	0,140477	-0,072253
T, сек.	0,154955	-0,927417	-0,147925	-0,282713	-0,052830
RR, сек.	-0,599627	-0,717035	0,042596	0,247799	0,200873

Следующий этап работы заключается в нахождении факторных значений индивидуумов по формуле:

$$\mathbf{F} = \mathbf{A}^T \mathbf{R}^{-1} \tilde{\mathbf{Z}}, \quad (9)$$

где \mathbf{A} – матрица весовых нагрузок, полученная итерационным методом главных факторов и подвергнутая варимакс-вращению, \mathbf{R} – корреляционная матрица стандартизованных данных \mathbf{Z} , $\tilde{\mathbf{Z}}$ – матрица нормированных данных, \mathbf{F} – матрица значений искомым факторных значений размерности $g \times n$.

На рис. 1. представлена факторная диаграмма здоровых индивидуумов, а на рис. 2 – факторная диаграмма индивидуумов, имеющих патологию сердечно-сосудистой системы. Факторные значения здоровых индивидуумов не превышают границы 2-й дисперсии. Факторные диаграммы индивидуумов с кардиопатологией имеют характерный вид. В этом случае факторные значения 2-го и

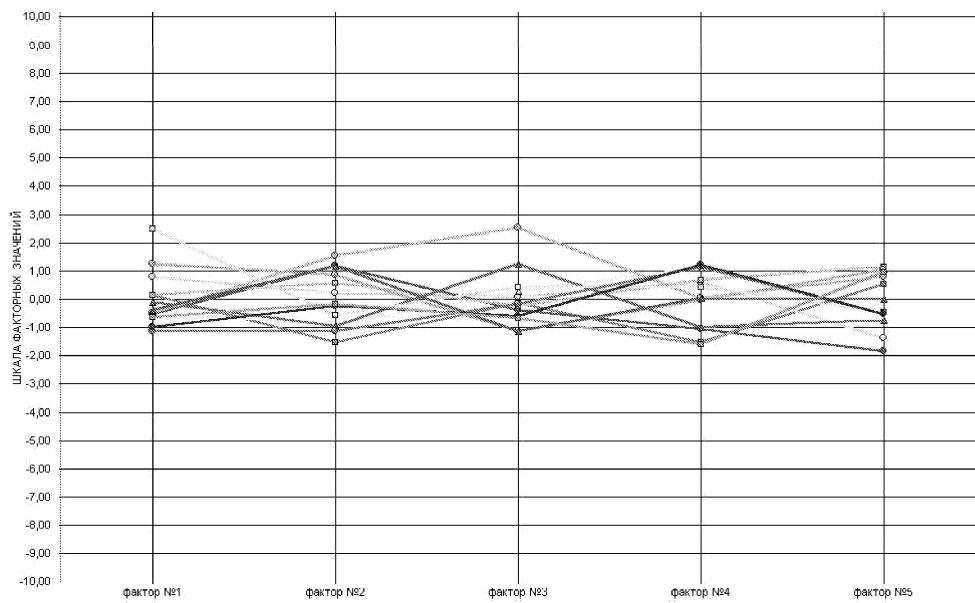


Рис. 1. Факторная диаграмма относительно здоровых индивидуумов.

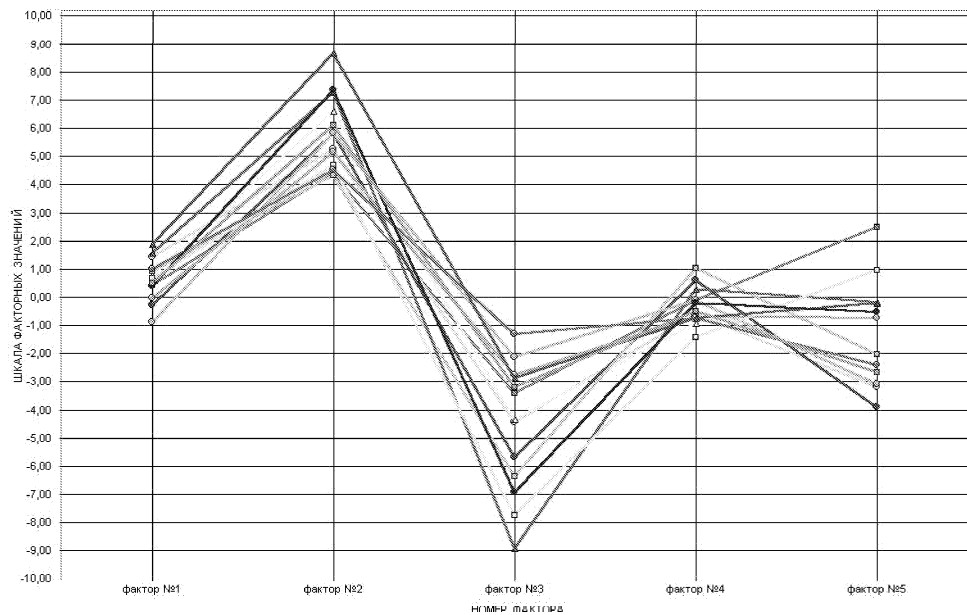


Рис. 2. Факторная диаграмма индивидуумов с патологией.

3-го факторов превышают границы 2-й дисперсии и достигают значений больше 8, разброс факторных значений 5-го фактора носит хаотичный характер и превышает норму на одну единицу. Одновременный анализ оптимальной матрицы весовых нагрузок и факторной диаграммы индивидуумов, имеющих патологию, показывает, что два фактора (2-й и 3-й) характеризуют данную сердечно-сосудистую патологию. Каждый из этих факторов имеет свое параметрическое наполнение и соответственно описывает определенный патологический процесс. Следовательно, по этим факторам можно не только проводить диагностику, но они являются определяющими при коррекции патологического состояния.

В результате анализа полученных Q-структур установлено, что положение сердца в грудной клетке является ведущим показателем, определяющим проекцию потенциала действия предсердий и желудочков на три плоскости: фронтальную, горизонтальную, сагиттальную.

Из всех рассмотренных и оцененных с помощью факторного анализа количественных показателей ЭКГ наиболее доказательными являются: отклонение электрической оси QRS, длины зубцов P, Q, R, которые имеют свои четкие закономерности проявления. Полученные корреляционные взаимосвязи между параметрами ЭКГ, факторные структуры и факторные диаграммы для лиц с различными антропометрическими характеристиками позволяют повысить чувствительность и специфичность выявления ряда патологических состояний по данным ЭКГ-исследования, например постановку диагнозов: «гипертрофия левого желудочка», «гипертрофия правого желудочка».

ЛИТЕРАТУРА

1. Иберла И. *Факторный анализ*. М.: Статистика, 1980. 399 с.
2. Гольдяпин В.В., Топчий В.А., Терентьев С.А. *Факторная модель гомеостаза для диагностики митрального стеноза различной степени // Микросенсорика*. Сб. трудов НИИ ИСМЭ СО РАН. 2000. С. 95-101.
3. Харман Г. *Современный факторный анализ*. М.: Статистика, 1972. 486 с.