

## СПИСОК ЛИТЕРАТУРЫ

1. Яковлев В.М., Нечаева Г.И., Готов А.В. Терминология, определение с позиции, классификация врожденной дисплазии соединительной ткани // Врожденная дисплазия соединительной ткани: тезисы регионального симпозиума. – Омск, 1990. – С. 3–5.
2. Яковлев В.М., Готов А.В., Ягода А.В. Иммунопатологические синдромы при наследственной дисплазии соединительной ткани. – Ставрополь, 2005.
3. Шараев П.Н., Иванов В.Г., Гаврилов А.Л. Методы лабораторного исследования показателей обмена коллагена в биологических жидкостях: Информационное письмо для врачей клинической лабораторной диагностики. – Ижевск, 2003. – С. 19.
4. Иберла К. Факторный анализ. – М.: Статистика, 1980.
5. Гольяпин В.В. Вычислительные аспекты метода минимальных остатков при разрешении варианта Хейвуда. // Сибирский журнал промышленной математики. – 2005. – Том VIII, №3(23). – С. 145–151.
6. Харман Г. Современный факторный анализ. – М.: Статистика, 1972.

# ФАКТОРИЗОВАННЫЕ МЕТОДЫ РАСПОЗНАВАНИЯ ОБРАЗОВ В КОМПЬЮТЕРНОЙ ДИАГНОСТИКЕ ДИСПЛАЗИИ СОЕДИНИТЕЛЬНОЙ ТКАНИ

Готов А.В., Гольяпин В.В.,  
Лобачев А.И.

Омский государственный университет  
им. Ф.М. Достоевского, Омск

## Введение

В развитии синдрома соединительнотканной дисплазии (СТД) принимают участие многие факторы. Как правило, в них задействованы эндогенные и экзогенные механизмы.

Увеличение числа случаев СТД, наблюдающееся в настоящее время, связано с патогенными воздействиями в онтогенезе, связанными с ухудшением экологической обстановки. Так же немаловажными факторами являются плохое питание и стрессы.

На сегодняшний день одной из задач современной медицины является исследование дисплазии соединительной ткани. Это состояние, которое зависит от молекулярных, генетически детерминированных дефектов биосинтеза компонентов соединительной ткани (каллогена, эластина, фибрина).

В связи с увеличением случаев СТД и, соответственно, с возросшей актуальностью данной проблемы, основной целью стала реализация факторной модели средовых воздействий и создание программного комплекса на базе факторизованных методов распознавания образов, решающего диагностическую задачу, посвященную дисплазии соединительной ткани подростков. В целях проведения соответствующего научного исследования специалистами была составлена анкета для опроса детей школьного возраста, характеризующая социальный статус индивидуумов с СТД и без него.

## Факторизованные методы распознавания образов

Прежде чем приступать к изложению факторизованных методов распознавания образов, сформулируем основные понятия факторного анализа. Одной из основных целей факторного анализа является извлечение на поверхность фактора, который бы наиболее точно позволил воспроизвести наблюдаемые корреляции.

Для любого факторного исследования исходные данные записываются в виде матрицы  $Y = \{y_{ij}\}$ , где индекс  $i = 1, 2, \dots, m$  относится к переменным (показателям), а индекс  $j = 1, 2, \dots, n$  к объектам (индивидуумам). Если исходные данные распределены по нормальному закону распределения, то имеется возможность проведения стандартизации матрицы  $Y$ . Тогда для стандартизованных данных имеет место соотношение:

$$R = \frac{1}{n-1} ZZ^T, \quad (1)$$

где  $Z = \{z_{ij}\}$  – матрица стандартизованных исходных данных;  $R$  – корреляционная матрица.

Целью любого метода факторного анализа является представление величины  $z_{ij}$ , то есть

элемента матрицы  $Z$ , в виде линейной комбинации нескольких гипотетических переменных, или факторов. Т.е. значение  $z_{ij}$  может быть выражено в виде линейной комбинации факторов:

$$z_{ij} = a_{i1}f_{1j} + a_{i2}f_{2j} + \dots + a_{ir}f_{rj}, \quad (2)$$

где  $a_{ir}$  – это весовые нагрузки факторов, а  $f_{ij} - f_{rj}$  – факторные значения факторов  $y$ -го индивидуума. Равенство (2) выражает основную модель факторного анализа.

Отдельные наблюдаемые значения являются линейными комбинациями гипотетических, ненаблюдаемых, или скрытых переменных, называемых факторами, которые не могут быть обнаружены в процессе наблюде-

ния. Матрица  $Z$  представляет собой произведение двух матриц  $A$  и  $F$ . При этом матрица  $A$  отражает связи переменных с факторами, а  $F$  описывает отдельные индивидуумы. Если предполагается отсутствие корреляции между факторами, то, подставив формулу (2) в формулу (1), получаем фундаментальную теорему факторного анализа для ортогональных факторов:

$$R = AA^T. \quad (3)$$

Фундаментальная теорема утверждает, что корреляционная матрица может быть воспроизведена с помощью факторного отображения и корреляций между факторами [1].

В методе главных факторов в формуле (3) используется редуцированная матрица  $R_h$ . По диагонали этой матрицы стоят общности, которые были найдены методом минимальных остатков [2].

Метод главных факторов заключается в поиске главных компонент редуцированной корреляционной матрицы. Проблема сводится к классической задаче нахождения собственных значений и собственных векторов симметрической матрицы [3, 4, 5].

Матрицу весовых нагрузок факторов будем находить по следующей формуле:

$$A = U\Lambda^{1/2},$$

где  $U$  и  $\Lambda$  – матрицы собственных векторов и собственных значений матрицы  $R_h$  соответственно [6].

Далее возникает проблема количества выделяемых факторов, которую можно решать различными методами. В данной работе был использован метод, основанный на критерии «каменистой осыпи» [1].

Найденные матрицы весовых нагрузок факторов для каждого факторного пространства, позволяют реализовать подходящие методы распознавания образов в этих пространствах.

Для построения алгоритмов распознавания используются вероятностные методы распознавания, основанные на теории статистических решений. В общем случае применение вероят-

ностных методов распознавания предусматривает наличие вероятностных зависимостей между признаками распознаваемых объектов и классами, к которым эти объекты относятся [7]. В данном случае используется следующее вероятностное решение:

**Поскольку в факторном пространстве однородные факторные значения не превышают трех, то достаточно проверить, попадают ли вычисленные факторные значения объекта исследования в интервал  $[-3; 3]$ . В этом случае исследуемый объект относится к данному факторному пространству с большей долей вероятности.**

При этом факторные значения высчитываются по следующей формуле:

$$F = \Lambda^{-1} A^T Z, \quad (5)$$

где  $\Lambda$  – диагональная матрица собственных значений матрицы  $R$  [8].

В качестве меры близости в детерминированных системах распознавания часто используется среднеквадратичное расстояние между данным объектом и совокупностью объектов

[7]. Так как данное факторное пространство имеет форму сферы с центром в точке ноль, то расстояние до центра рассчитывалось как корень из суммы квадратов факторных значений  $i$ -го индивидуума в  $j$ -м факторном пространстве:

$$d_{ij} = \sqrt{f_{(j)1i}^2 + f_{(j)2i}^2 \dots f_{(j)ri}^2}, \quad (6)$$

где  $d_{ij}$  – эвклидово расстояние  $i$ -го индивидуума в  $j$ -м факторном пространстве, а  $f_{(j)ki}$  есть  $k$ -факторное значение  $i$ -го индивидуума в

$j$ -м факторном пространстве. В данном случае используется следующее вероятностное решение:

*Исследуемый объект относится к тому факторному пространству с большей долей вероятности, до которого факторизованное расстояние, вычисляемое по формуле (6), меньше.*

#### **Алгоритмы факторизованных методов распознавания образов**

Алгоритм нахождения матрицы факторных значений

1. Формирование таблицы с ответами респондентов.
2. Преобразование ответов респондентов в шкалы по типу «обратных школьных оценок».
3. Ранжирование данных.
4. Стандартизация исходных данных.
5. Вычисление корреляционной матрицы по формуле (1).
6. Нахождение матрицы весовых нагрузок по формуле (4).

#### **Алгоритм реализации факторизованных методов распознавания образов для качественных показателей**

1. Формирование объекта исследования:
  - создается вектор ответов для исследуемого объекта по установленной таблице опросов;
  - выбирается соответствие имеющихся ответов с базой шкалированных и ранжированных данных;
  - формируется вектор ранжированных данных конкретного объекта;
  - осуществляется стандартизация ранжированного вектора.
2. Вычисление факторных значений по формуле (5).
3. Проверка вложенности факторных значений соответствующего объекта в  $[-3; 3]$ .
4. Расчет факторизованных евклидовых расстояний по формуле (6).
5. Принятие решения на базе соответствующих вероятностных правил.

#### **Результаты и выводы**

Программная реализация вышеописанных алгоритмов позволила создать диагностико-исследовательский программный комплекс, основанный на факторизованных методах распознавания, и провести факторное исследование.

В результате проведенного факторного исследования выделены следующие факторы: загрязнения атмосферного воздуха промышленными предприятиями и автотранспортом,

радиационное и нерадиационное излучение, бытовые интоксикации, бытовые вредности, эпидермальная вредность, личностный фактор, факторы реализации, фактор превентивный по отношению к СТД. У пациентов с СТД оказались значимыми факторы окружающей среды обитания, которые оказали неблагоприятное воздействие на состояние их здоровья.

Полученные данные позволяют говорить о том, что все лица молодого возраста имеют отклонения в состоянии здоровья, связанные с воздействием неблагоприятных факторов окружающей среды обитания.

Данные изменения носят односторонний, но неоднозначный характер в исследовании групп. У лиц с СТД, по сравнению со здоровыми, значимых неблагоприятных факторов больше, чем в контрольной группе. И характер их воздействия затрагивает среду обитания как дома, так и в учебном процессе.

#### **СПИСОК ЛИТЕРАТУРЫ**

1. Иберла К. Факторный анализ / Иберла К.; пер. с нем. В.М. Ивановой. – М.: Статистика, 1980. – 397 с.
2. Гольяпин В.В. Вычислительные аспекты метода минимальных остатков при разрешении варианта Хейвуда // Сибирский журнал индустриальной математики. – 2005. – Т. VIII, №3 (23). – С. 121-129.
3. Thurstone L.L. Multiple factor analysis / L.L. Thurstone // 6th. ed. – Chicago, 1961.
4. Zurmühl R. Matrizen und ihre technischen Anwendungen / R. Zurmühl. – Berlin, 1964.
5. Wilkinson J.H. The algebraic eigenvalue problem / J.H. Wilkinson. – Oxford, 1965; Русский пер. Дж.Х. Уилкинсон Алгебраическая проблема собственных значений / Дж.Х. Уилкинсон. – М.: Наука, 1970. – 565 с.
6. Harman H. Modern factor analysis. / H. Harman. – Chicago, 1967; Русский пер. Г. Харман Современный факторный анализ / Г. Харман. – М.: Статистика, 1972. – 483 с.
7. Горелик А.Л. Методы распознавания, издание второе, переработанное и дополненное / А.Л. Горелик, В.А. Скрипкин. – М.: Высшая школа, 1984. – 208 с.
8. Гольяпин В.В. Использование свойств ортогональности факторов в вероятностной задаче распознавания // Сибирский журнал индустриальной математики. – 2007. – Т. X, №2 (30). – С. 53–63.