

более часто используемых конечных элементов в виде параллелепипеда и тетраэдра, позволяющий осуществить целесообразный выбор вида конфигурации конечного элемента для моделирования НДС кортикального слоя костных тканей.

Для расчета величины деформации костной ткани под действием сжимающего усилия из кортикального слоя кости вырезан элемент в виде балки квадратного сечения. Один конец балки имеет жесткую заделку, а к другому приложено сжимающее давление.

Полученные результаты показали, что для моделирования НДС костной ткани можно использовать как тетраэдральный конечный элемент, так и в виде параллелепипеда. При использовании в качестве конечного элемента параллелепипеда мы имеем ряд преимуществ. Во-первых, это разбиение конструкции на регулируемые конечные элементы. Во-вторых, для узлов не задаются координаты, а выполняется их нумерация. В-третьих, матрица жесткости для параллелепипеда зависит только от его размеров и может храниться в памяти ЭВМ, что позволит значительно сократить время расчетов. Основной трудностью использования тетраэдрального элемента является задание для каждого узла пространственных координат: это трудоемкий процесс, требующий больше времени для анализа. Однако костная ткань имеет сложное геометрическое и физическое строение, для ее моделирования удобнее применять тетраэдральные элементы. Поэтому при разработке программы для расчета прочностных характеристик костных тканей предлагается использовать конечный элемент в виде тетраэдра.

#### ЛИТЕРАТУРА

1. Проблемы прочности в биомеханике. / под ред. *И.Ф. Образцова*. – М.: Наука, 1988.
2. *Бегун П.И., Шукейло Ю.А.* Биомеханика: учебник для вузов. – СПб.: Политехника, 2000.
3. *Мюллер М.Е, Альговер М.* Руководство по внутреннему остеосинтезу. – М.: Наука, 1996.

## **НЕКОТОРЫЕ ОСОБЕННОСТИ ПРОЦЕДУРЫ КЛАСТЕРИЗАЦИИ МЕДИКО-ЭКОЛОГИЧЕСКИХ ДАННЫХ**

**Е.В. Черныш**

(Дальневосточный государственный университет, Владивосток)

В настоящее время многие медицинские и экологические исследования, связанные с моделированием изучаемых процессов, основаны на использовании процедуры кластеризации, которая включает различные математические и статистические методы обработки информации с целью подготовки и структурирования исходных данных. Кластеризация позво-

ляет разбить исходное множество объектов на заданное или заранее неизвестное число кластеров (классов) однородных объектов по некоторому математическому критерию. Кластер характеризуется плотностью (плотность объектов внутри кластера выше, чем вне его), дисперсией (чем больше дисперсия расстояния от центра до остальных объектов кластера, тем данный кластер более разрежен), отделимостью от других кластеров, а также формой, – например, кластер может иметь сложную цепочечную структуру, быть в виде сегмента или иметь очертания поверхности как гиперсфера или эллипсоид.

С математической точки зрения задачей кластеризации данных является выбор метрик, на основе которых можно определить меру сходства объектов, расстояние между объектом и классом, расстояние между целыми группами объектов. Центр кластера есть среднее геометрическое место точек в пространстве признаков. Расстояние между объектом и целым классом рассматривается как расстояние от объекта до центра данного кластера. Выбор меры сходства объектов решающим образом влияет на окончательный вариант разбиения множества объектов на группы при заданном алгоритме разбиения. В свою очередь, выбор меры межкластерного расстояния (например, по центрам кластеров, по ближайшим или по самым дальним объектам кластеров) влияет на вид получаемых с помощью классификации геометрических подмножеств объектов в пространстве признаков. Существенным элементом кластерного анализа является определение метода группировки объектов.

Задача классификации усложняется, поскольку объекты являются многомерными, т.е. описываются несколькими параметрами (признаками), среди которых могут встречаться данные нечислового характера. К тому же различные признаки часто имеют разные диапазоны изменения или разные размерности. Поэтому на начальном этапе кластеризации необходима нормировка исходных данных: на среднеквадратичное отклонение, на выборочное среднее или на фоновое значение. Технически удобно использовать данные, нормированные несколькими способами, что позволит выявить наиболее важные переменные и сузить пространство рассматриваемых признаков. Как правило, геохимические данные и данные по заболеваемости относятся к административным районам большой протяженности и поэтому усредняются (берется среднее геометрическое). При таком усреднении целесообразно применение весовых коэффициентов, пропорциональных численности населения, проживающего в соответствующем районе.

В биомедицинских исследованиях наиболее часто применяется иерархическая классификация данных, где первоначально все объекты рассматриваются как отдельные кластеры из одного элемента. А затем, используя стандартные метрики (евклидову, «взвешенную» евклидову, Махаланобиса, Хемминга и др.), последовательно, шаг за шагом, кластеры

объединяются между собой.

Для классификации территории по заболеваемости населения или по геохимическим данным окружающей среды традиционно за основу берется итеративный метод группировки  $k$ -средних, при котором задается начальное (искусственное) разбиение множества объектов на классы и определяется некоторый математический критерий качества автоматической классификации, а затем объекты перераспределяются из кластера в кластер, пока значение критерия не станет улучшаться. В качестве такого критерия задается конкретное число итераций либо достигается стабильность классификации (начиная с некоторого шага кластеры перестают изменяться). В задачах обработки экологических данных обычно выбирается евклидова метрика, нередко используется так называемое «угловое расстояние», имеющее геометрический смысл косинуса угла между двумя векторами в пространстве признаков. Алгоритмы, основанные на базе  $k$ -средних, отличаются от других методов простотой реализации и практической надежностью.

Многообразие алгоритмов кластерного анализа обусловлено также множеством различных критериев оптимальности классификации, которые могут быть составлены совершенно разными способами. Наиболее часто такие критерии основываются на мерах внутрикластерного и межкластерного разброса, на отношениях показателей плотности кластеров к расстоянию между ними, на отношениях общей дисперсии данных к сумме внутриклассовых дисперсий и дисперсии центров кластеров и т.д.

Такие алгоритмы процедуры кластеризации характеризуются определенной трудоемкостью и требуют ресурсов высокопроизводительных компьютеров. Разнообразные алгоритмы классификации входят в состав многих современных пакетов прикладных программ для обработки многомерных данных.

Представленная работа выполнена в рамках гранта НШ-9004.2006.