

АНАЛИЗ МЕТОДОВ РАСПОЗНАВАНИЯ СПОСОБА ТРАВМИРУЮЩЕГО ВОЗДЕЙСТВИЯ ПО ПАРАМЕТРАМ РАЗРУШЕНИЙ ТРУБЧАТЫХ КОСТЕЙ

В. А. Голуб, Н. В. Огаркова

Воронежский государственный университет

В работе рассматриваются различные методы распознавания способа травмирующего воздействия на основе кластерного распознавания образов для последовательной двухэтапной процедуры анализа параметров разрушения костей.

Приведены результаты, полученные при применении различных методов, определен оптимальный кластерный метод распознавания способа травмирующего воздействия.

ВВЕДЕНИЕ

Одной из актуальных задач, решаемых в судебной медицине, является определение способа травматического воздействия при автомобильной травме (травме в результате дорожно-транспортного происшествия). Такими воздействиями могут быть удар, когда автомобиль сбивает пострадавшего, либо давление в случае переезда пострадавшего транспортным средством. Возможны сочетания разрушающих воздействий, когда повторные травматические воздействия обусловлены вторичным переездом пострадавшего. В судебно-медицинской практике способ травматического воздействия может быть определен по признакам и характеру разрушения длинных трубчатых костей.

Наиболее значимые для судебно-медицинского заключения о способе травмирующего воздействия параметры были определены на основе результатов исследований повреждений костей, выполненных в частности в работах [1–5]. Полный список этих параметров приведен в работах [7,8]. Так как почти все эти параметры представляются случайными величинами, то в качестве их математических характеристик используются числовые выборочные характеристики.

МАТЕМАТИЧЕСКАЯ ПОСТАНОВКА ЗАДАЧИ

Для решения задачи определения вида травмирующего воздействия могут быть использованы методы теории распознавания образов, когда совокупность N анализируемых параметров, характеризующих повреждение кости

$x = (x_1, x_2, \dots, x_N)$ рассматривается как вектор образа, который должен быть отнесен к одному из четырех возможных классов: ω_1 — удар, ω_2 — давление, ω_3 — удар с последующим давлением, ω_4 — давление с последующим давлением, где удар с последующим давлением и давление с последующим давлением характеризуют повторную травматизацию, возникающую при переезде пострадавшего транспортным средством после наезда на него.

АЛГОРИТМ РАСПОЗНАВАНИЯ

Как показано в [8], наилучшим образом смоделировать действия судебно-медицинского эксперта можно с помощью последовательной процедуры анализа параметров разрушения костей, включающую два этапа. На первом этапе осуществляется классификация на два класса ω_1 и ω_2 — удар или давление, а на втором этапе, на основании анализа наличия признаков повторной травматизации, также осуществляется классификация на два класса — либо удар (ω_1) или удар с последующим давлением (ω_3), либо давление (ω_2) или давление с последующим давлением (ω_4).

Указанная классификация может быть эффективно осуществлена на основе кластерного распознавания образов [6]. Такой подход предполагает объединение эталонных векторов образов в классы — кластеры и отнесение распознаваемого вектора образа к тому или иному классу на основании значения функции расстояния «образ-кластер», являющейся количественной мерой сходства анализируемого и эталонных образов.

Для процедуры распознавания (классификации) могут быть использованы различные

методы. Для их описания введем некоторые понятия. Поскольку на каждом этапе выполнения рассмотренного выше алгоритма распознавания, осуществляется классификация на два класса, то будем считать, что количество классов M равно двум ($M=2$), (это является существенным достоинством предлагаемого двухэтапного алгоритма [8], обеспечивающим упрощение задачи и, следовательно, более быструю работу компьютерной программы, реализующей такой алгоритм). Пусть дана выборка образов с известной классификацией $\{s_1, s_2, \dots, s_N\}$, причем предполагается, что каждый образ входит в один из классов $\omega_j, j=1, 2$; S_j — множество эталонных образов, относящихся к j -ому классу (кластеру) — обучающая выборка.

Метод распознавания по *расстоянию до центра кластера* [6]. Для каждого кластера сначала определяется его центр $\mathbf{m}_j = \frac{1}{N_j} \sum_{\mathbf{x} \in S_j} \mathbf{x}$, $j=1, 2$. Затем рассчитываются расстояния между вектором \mathbf{x} и центром каждого кластера $l_j = D(\mathbf{m}_j, \mathbf{x})$, $j=1, 2$. Вектор \mathbf{x} зачисляется в тот класс, расстояние до центра кластера которого минимально, причем вероятность того, что вектор \mathbf{x} относится к первому классу будет определяться формулой $P_1 = \frac{l_2}{l_1 + l_2}$, а вероятность отнесения вектора \mathbf{x} ко второму классу — $P_2 = \frac{l_1}{l_1 + l_2}$.

Метод распознавания по *расстоянию до ближайшего соседа* [6] (БС-метод). Классифицируемый образ \mathbf{x} относится тому классу, к которому принадлежит его ближайший сосед, причем образ $s_i \in \{s_1, s_2, \dots, s_N\}$ называется ближайшим соседом образа \mathbf{x} , если

$$D(s_i, \mathbf{x}) = \min_j \{D(s_j, \mathbf{x})\}, \quad j=1, 2,$$

где D — любая из применимых метрик. В этом случае вероятности отнесения вектора \mathbf{x} к тому или иному кластеру будут определяться также как при использовании метода распознавания по расстоянию до центра кластера. Расстояния l_1 и l_2 в этом случае будут определяться как расстояния до ближайших соседей из первого и второго кластеров соответственно.

Метод распознавания по «*q* ближайших соседей» [6] (*q*-БС-метод). Определяются q ближайших соседей для вектора \mathbf{x} и он зачисляется в тот класс, к которому относится наибольшее число образов, входящих в эту группу. Если

q_1 — количество образов, которые относятся к классу ω_1 , а q_2 — количество образов, которые относятся к классу ω_2 , причем $q_1 + q_2 = q$, то вероятность отнесения вектора \mathbf{x} к классу ω_1 определяется формулой $P_1 = q_1/q$, а к классу ω_2 — $P_2 = q_2/q$.

Метод распознавания «*q*-БС» может быть модифицирован с целью потенциального улучшения характеристик распознавания. Рассмотрим два возможных варианта этого метода.

Вариант 1 q-БС-метода (q-БС-1). Обозначим через Q группу из q ближайших соседей, причем q_1 — количество образов которые относятся к классу ω_1 и образуют множество Q_1 , а q_2 из них относятся к классу ω_2 и образуют множество Q_2 . Для каждого класса определяется среднее расстояние по формуле: $l_j = \sum_{s_i \in Q_j} D(s_i, \mathbf{x}) / q_j$,

$j=1, 2$. При этом вектор \mathbf{x} заносится в тот класс, где среднее расстояние l_j окажется меньше. При этом вероятность отнесения вектора \mathbf{x} в класс ω_1 будет определяться формулой $P_1 = l_2/l$, а вероятность отнесения вектора \mathbf{x} к классу ω_2 — $P_2 = l_1/l$, где $l = l_1 + l_2$.

Вариант 2 q-БС-метода (q-БС-2). Определяются q ближайших соседей для из классов ω_1 и ω_2 . Обозначим множества соседей Q_1 и Q_2 соответственно. Для каждого из этих множеств определяется суммарное расстояние до вектора \mathbf{x} по формуле $l_j = \sum_{s_i \in Q_j} D(s_i, \mathbf{x})$. Вектор \mathbf{x} зачисляется в тот класс, где l_j окажется меньше. При этом вероятность отнесения вектора в класс ω_1 будет определяться формулой $P_1 = l_2/l$, а вероятность отнесения к классу ω_2 — $P_2 = l_1/l$, где $l = l_1 + l_2$.

В качестве меры сходства наиболее часто используются следующие функции [6]:

— евклидово расстояние

$$D = \|\mathbf{x} - \mathbf{z}\| = \sqrt{\sum_{k=1}^m (x_k - z_k)^2}; \quad (1)$$

— нормированное евклидово

$$d_{ij} = \sqrt{\sum_{k=1}^m \frac{(x_{ik} - x_{jk})^2}{S_k^2}}; \quad (2)$$

— сумма квадратов

$$d_{ij} = \frac{1}{n} \sum_{k=1}^m (x_{ik} - x_{jk})^2; \quad (3)$$

— взвешенная сумма квадратов

$$d_{ij} = \frac{1}{n} \sum_{k=1}^m w_k^2 (x_{ik} - x_{jk})^2; \quad (4)$$

— расстояние Махаланобиса

$$D = (\mathbf{x} - \mathbf{m})' \mathbf{C}^{-1} (\mathbf{x} - \mathbf{m}), \quad (5)$$

где \mathbf{C} — ковариационная матрица совокупности образов кластера, \mathbf{m} — вектор средних значений.

Решение вопроса, применение какой из этих функций дает наилучшие результаты распознавания, требует специального анализа для каждой конкретной задачи.

АЛГОРИТМ АНАЛИЗА ЭФФЕКТИВНОСТИ МЕТОДОВ РАСПОЗНАВАНИЯ

Эффективность того или иного метода, используемого в рамках кластерного распознавания образов, может оцениваться величиной вероятности правильной классификации, которая, в свою очередь, может быть определена

— либо по количеству правильно классифицированных образов ($K_{\text{пр}}$), отнесенному к их общему числу (K) ($P_1 = K_{\text{пр}} / K$);

— либо по среднему значению оценки вероятности правильной классификации образов, определяемой как $P_2 = \frac{1}{K} \sum_{i=1}^K P_i'$, где P_i' — вероятность правильной классификации i -го образа.

Так как на начальном этапе нет никаких данных, кроме обучающих выборок, то для анализа эффективности методов в обоих случаях может быть применен следующий алгоритм: один из образов, классификация которого известна, исключается из обучающей выборки и затем осуществляется его классификация на основе всех оставшихся в выборке образов.

При оценке эффективности метода распознавания по первому критерию определяется количество правильно классифицированных образов, и затем эффективность метода распознавания оценивается отношением правильно классифицированных образов к их общему количеству.

При оценке эффективности по второму критерию для каждого из анализируемых векторов образов определяется вероятность правильной классификации, а затем определяется их среднее значение.

РЕЗУЛЬТАТЫ

Описанная выше процедура анализа эффективности методов распознавания была исполь-

зована применительно к задаче определения вида травмирующего воздействия. В качестве обучающих использовались выборки, сформированные по данным экспериментов, в которых изучались изменения параметров разрушения длинных трубчатых костей (большой берцовой и бедренной) для различных видов травмирующего воздействия: ударе и давлении, как при отсутствии, так и при наличии повторной травматизации. На каждом этапе классификация осуществлялась на два класса в соответствии с двухэтапным алгоритмом, описанным выше. При этом для распознавания использовалось двадцать шесть параметров, характеризующих разрушение кости, то есть размерность вектора образа составляла $N = 26$. В качестве меры сходства (функции расстояния) использовалось евклидово расстояние.

В табл. 1—4 приведены результаты вычислений вероятности правильной классификации вида травмирующего воздействия (P_1) для различных методов классификации образов в рамках кластерного метода распознавания. В первом столбце представлено значение параметра q — количество ближайших к анализируемому образам (соседей). В остальных столбцах представлены результаты, полученные при использовании различных стратегий:

1 — метод распознавания по расстоянию до центра кластера;

2 — метод распознавания по расстоянию до ближайшего соседа;

3 — метод распознавания по расстоянию до q ближайших соседей (q -БС);

4 — вариант 1 q -БС-метода (q -БС-1);

5 — вариант 2 q -БС-метода (q -БС-2).

ВЫВОДЫ

1. Результаты вычисления эффективности методов распознавания по количеству правильно классифицированных образов (P_1) и по среднему значению оценки вероятности правильной классификации образов (P_2) отличаются незначительно. Поэтому для дальнейшего анализа целесообразно использовать оценку эффективности методов распознавания по количеству правильно классифицированных образов (P_1).

2. В методах распознавания по q ближайшим соседям методе, брать число «соседей» $q > 5$ нецелесообразно, так как значение вероятности правильной классификации при этом не увели-

Таблица 1

Значения вероятностей правильной классификации вида травмирующего воздействия для бедренной кости без учета повторной травматизации

<i>q</i>	1	2	3	4	5
0	0.722	—	—	—	—
1	—	0.759	—	—	0.759
2	—	—	0.759	0.759	0.772
3	—	—	0.734	0.797	0.747
4	—	—	0.778	0.722	0.747
5	—	—	0.759	0.696	0.734
6	—	—	0.772	0.722	0.747
7	—	—	0.797	0.772	0.747
8	—	—	0.759	0.785	0.747
9	—	—	0.772	0.785	0.722
10	—	—	0.741	0.722	0.722

Таблица 2

Значения вероятностей правильной классификации вида травмирующего воздействия для большой берцовой кости без учета повторной травматизации

<i>q</i>	1	2	3	4	5
0	0.308	—	—	—	—
1	—	0.673	—	—	0.673
2	—	—	0.663	0.673	0.731
3	—	—	0.750	0.635	0.750
4	—	—	0.769	0.615	0.750
5	—	—	0.769	0.596	0.731
6	—	—	0.760	0.500	0.712
7	—	—	0.769	0.385	0.731
8	—	—	0.750	0.442	0.769
9	—	—	0.750	0.442	0.769
10	—	—	0.760	0.442	0.769

Таблица 3

Значения вероятностей правильной классификации вида травмирующего воздействия для бедренной кости с учетом повторной травматизации

<i>q</i>	1	2	3	4	5
0	0.228	—	—	—	—
1	—	0.165	—	—	0.494
2	—	—	0.620	0.582	0.494
3	—	—	0.139	0.190	0.494
4	—	—	0.171	0.177	0.456
5	—	—	0.089	0.215	0.456
6	—	—	0.139	0.190	0.456
7	—	—	0.101	0.228	0.456
8	—	—	0.082	0.190	0.456
9	—	—	0.139	0.165	0.000
10	—	—	0.114	0.139	0.000

Таблица 4

Значения вероятностей правильной классификации вида травмирующего воздействия для большой берцовой кости с учетом повторной травматизации

<i>q</i>	1	2	3	4	5
0	0.596	—	—	—	—
1	—	0.481	—	—	0.135
2	—	—	0.481	0.481	0.135
3	—	—	0.596	0.500	0.135
4	—	—	0.567	0.423	0.135
5	—	—	0.519	0.500	0.135
6	—	—	0.519	0.519	0.135
7	—	—	0.538	0.519	0.135
8	—	—	0.538	0.481	0.135
9	—	—	0.558	0.519	0.135
10	—	—	0.548	0.519	0.135

чивается, а сложность вычислений существенно возрастает.

3. Правильное распознавание вида травмирующего воздействия без анализа наличия повторной травматизации осуществляется с достаточно высокой вероятностью ($\sim 0,75$) как по разрушению бедренной, так и по разрушению большой берцовой кости, но в первом случае вероятность правильного распознавания несколько выше.

4. Для определения вида травмирующего воздействия без анализа наличия повторной травматизации по бедренной кости наилучшие результаты достигаются при использовании метода распознавания по q ближайшим соседям, другие методы дают практически такие же хорошие результаты.

5. При определении вида травмирующего воздействия без анализа наличия повторной травматизации наиболее высокие значения вероятности правильного распознавания также достигаются при использовании методов распознавания q -БС и варианта 2 q -БС-метода (q -БС-2). Использование метода распознавания по расстоянию до центра кластера дает существенно худшие результаты.

6. Вероятность правильного определения вида травмирующего воздействия с анализом наличия повторной травматизации при использовании двухэтапной процедуры распознавания значительно меньше, чем без анализа наличия повторной травматизации — разброс значений вероятностей оказывается достаточно большим. При этом результаты, полученные по параметрам разрушения большой берцовой кости, лучше, чем по параметрам разрушения бедренной.

7. При использовании данных о разрушении бедренной кости наилучшие результаты по определению вида травмирующего воздействия с анализом наличия повторной травматизации достигаются при использовании метода распознавания q -БС и варианта 2 q -БС-метода (q -БС-2).

8. Наилучшие результаты по определению вида травмирующего воздействия по разрушению большой берцовой кости с анализом наличия повторной травматизации достигаются при использовании метода распознавания q -БС либо метода распознавания по расстоянию до центра кластера.

Таким образом, наилучшие результаты по определению вида травмирующего воздействия могут быть получены при использовании метода распознавания по q ближайшим соседям (q -БС-метод), при этом значение параметра q лучше брать в диапазоне от 2 до 5.

Полученные результаты могут быть использованы при разработке алгоритмов и программных продуктов, предназначенных для автоматизации распознавания способа травмирующего воздействия по параметрам разрушения трубчатых костей.

СПИСОК ЛИТЕРАТУРЫ

1. Диагностикум механизмов и морфологии переломов при тупой травме скелета. Т. 1. Механизмы и морфология переломов длинных трубчатых костей / В. И. Бахметьев, В. Н. Крюков, В. П. Новоселов и др. — 2-е изд. — Новосибирск: Наука. Сибирская издательская фирма РАН, 2002. — 166 с.
2. Крюков В.Н. Пути и методы решения проблемы установления последовательности повреждений // Тез. докл. III Всесоюз. съезда суд. медиков. — М., Одесса, 1988. — С. 75—76.
3. Семенников В.С. Диагностика механизмов травмы по характеру и особенностям переломов длинных трубчатых костей // Актуальные вопр. суд. медицины и экспертной практики: Тез. докл. Ростов-на-Дону, 1985. — С. 73—74
4. Солохин А.А., Абдукаримов Р.Х. Судебно-медицинская диагностика видов автомобильной травмы с применением математических методов и программных систем // Судебно-медицинская экспертиза, 1991, № 3. 10—12 с.
5. Янковский В.Э. К вопросу об определении механизмов образования переломов // Актуальные вопр. суд. медицины и экспертной практики. — Барнаул, 1988. — С. 11—17.
6. Ту Дж. Т., Гонсалес Р.К. Принципы распознавания образов / Пер. с англ. И. Б. Гуревича; Под ред. Ю. И. Журавлева. — М.: Мир, 1978. — 411 с.
7. Бахметьев В.И. Применение метода бутстрепа к анализу параметров морфологических признаков разрушения трубчатых костей / В. И. Бахметьев, В. А. Голуб, Г. А. Князев, Н. В. Огаркова // Вестн. Воронеж. Ун-та. Сер. Системный анализ и информационные технологии. — № 1.
8. Бахметьев В.И. Алгоритм распознавания способа травмирующего воздействия по параметрам разрушений длинных трубчатых костей / В. И. Бахметьев, В. А. Голуб, Г. А. Князев, Н. В. Огаркова // Системные проблемы надежности, качества, информационных и электронных технологий в инновационных проектах (Инноватика-2006) / Материалы Международной конференции и Российской научной школы. Часть 5., Т. 1. — М.: Радио и связь, 2006. — С. 101—104.