

РАЗДЕЛ IV. ИЗМЕРЕНИЕ, МОДЕЛИРОВАНИЕ И УПРАВЛЕНИЕ В ЭКОЛОГИИ, НАУКАХ О ЧЕЛОВЕКЕ И ОБЩЕСТВЕ

Средняя вероятность ошибочной классификации по всему набору столбцов имеет вид:

$$p = \frac{\sum_j p_j}{m} = \frac{m^2 - \sum_j \sum_k r_j(k)}{m^2}. \quad (5)$$

По формуле (5) средняя вероятность ошибочной классификации составит 0,507.

Следовательно, получили однородность классов – 0,493. Так как был выбран уровень однородности 45% (0,45), то классы являются однородными. В противном случае процесс выявления набора продолжают циклически с переходом на $(j+1)$ столбец.

В общем виде при рассмотрении комбинации из s столбцов для вычисления уровня однородности используют следующие формулы:

$$h_{(i_1, \dots, i_s)g} = H_{i_1, \dots, i_s} + H_g - H_{i_1, \dots, i_s g};$$
$$I_{i_1 \dots i_s} = \frac{\sum_{g \neq i_1 \dots i_s} h_{(i_1 \dots i_s)g}}{\sum_{g \neq i_1 \dots i_s} H_{(i_1 \dots i_s)g}}.$$

Процедура классификации будет завершена, как только достигнут желаемый уровень однородности.

Таким образом, в статье предложен энтропийный способ декомпозиции матрицы качественных оценок по номинальной шкале.

СПИСОК ЛИТЕРАТУРЫ

1. Икрамов, Х. Д. Численные методы линейной алгебры / Х. Д. Икрамов. – М.: Знание, 1987. – 46 с.
2. Икрамов, Х. Д. Разреженные матрицы / Х. Д. Икрамов. – М.: Знание, 1989. – 48 с.
3. Шатихин, Л. Г. Структурные матрицы и их применение для исследования систем / Л. Г. Шатихин. – М.: Машиностроение, 1991. – 253 с.
4. Эстербю, О. Прямые методы для разреженных матриц / О. Эстербю, З. Златев. – М.: Мир, 1987. – 118 с.
5. Бешелев, С. Д. Математико-статистические методы экспертных оценок / С. Д. Бешелев, Ф. Г. Гурвич. – М.: Наука, 1980. – 180 с.
6. Мёллер, Ф. Роль энтропии в номинальной классификации / Ф. Мёллер // Математика (в)? социологии. Моделирование и обработка информации / Под ред. А. Аганбегяна, Х. Блейлока, Ф. Бородкина, Р. Будона, В. Капекки. – М.: Мир, 1977. – 385 с.
7. Панкова, Л. А. Организация экспертизы и анализ экспертной информации / Л. А. Панкова, А. М. Петровский, М. В. Шнейдерман. – М.: Наука, 1984. – 120 с.
8. Шибанов, Г. П. Информационные технологии // Порядок формирования экспертных групп и проведения коллективной экспертизы. – 2003 - №12. – С. 120-122.

Аспирант **А.Б. Архипова** тел. 8-913-246-76-22, arhipova_ab@mail.ru - каф. Вычислительных систем и информационной безопасности АлтГТУ им. И. И. Ползунова. Профессор **В.М. Белов** тел. (383) 269-83-28, vmbelov@mail.ru - каф. БуИТ Сибирского государственного университета телекоммуникаций и информатики.

УДК 57.087

ВЫБОР ЗНАЧИМЫХ ДЛЯ ПРОГНОЗИРОВАНИЯ ВРЕМЕНИ ВЫЖИВАНИЯ ПОКАЗАТЕЛЕЙ ПАЦИЕНТА

Д.Ю. Стрункин

В статье рассматривается задача прогнозирования времени выживания пациента. Для эффективного решения данной задачи необходимо среди всех показателей пациента отобрать те, что значимо влияют на время выживания. Приводится методика отбора, основанная на корреляционном анализе и экспертных оценках.

Ключевые слова: корреляционный анализ, экспертные оценки, согласованность мнений, прогнозирование времени выживания, значимость показателей.

Введение

Одной из важных задач в работе специализированных медицинских учреждений является задача оценки состояния пациента и прогнозирования его времени выживания: будет ли пациент жив через год, три, пять или более лет после лечения. Необходи-

мость и важность прогноза времени выживания пациента обусловлена рядом причин. Во-первых, это необходимо для адекватного выбора программы лечения (особенно с точки зрения возможных побочных эффектов, связанных с применением рискованных схем лечения), а во-вторых – это

ВЫБОР ЗНАЧИМЫХ ДЛЯ ПРОГНОЗИРОВАНИЯ ВРЕМЕНИ ВЫЖИВАНИЯ ПОКАЗАТЕЛЕЙ ПАЦИЕНТА

может позволить учреждению правильно распределять финансовые ресурсы, направляя их более эффективно [1].

На сегодняшний день самым распространенным подходом к прогнозированию времени выживания пациентов является использование множительных оценок Каплана-Мейера (КМ). Оценки КМ могут быть построены для групп пациентов, которых лечили определенным методом. Они позволяют учесть цензурированность, свойственную медицинским данным, и определить среднюю для группы вероятность выживания в течение определенного количества лет [2].

Однако с точки зрения прогнозирования времени выживания конкретного пациента множительные оценки КМ обладают существенным недостатком. Поскольку они строятся для больших групп пациентов, зачастую характеризующихся различным состоянием здоровья, сопутствующими заболеваниями и т.д., то они исключают возможность учета индивидуальных особенностей пациента при прогнозировании, что существенно снижает точность прогноза. Поэтому большое значение имеет разработка новых методов прогнозирования времени выживания пациентов с учетом их индивидуальных особенностей.

История болезни пациента содержит информацию анамнеза, объективного исследования, показатели клинических и лабораторных исследований и т.д. В дальнейшем информацию, содержащуюся в истории болезни будем называть показателями пациента. При разработке новых методов прогнозирования времени выживания пациентов важнейшее значение имеет определение тех показателей пациентов, которые значимо влияют на время выживания.

Постановка задачи

Задача определения показателей пациента, значимо влияющих на время выживания, может быть формализована следующим образом.

Обозначим:

$X = (X_1, \dots, X_n)$ - показатели пациента, характеризующие состояние его здоровья;

$Y = (Y_1, \dots, Y_m)$ - свойства пациента, означающие, что пациент жив в течение определенного количества лет после лечения. Значения свойств Y зависят от значений показателей X .

Показатели X и зависящие от них свойства Y являются переменными одного из трех типов:

- Дихотомические, которые могут принимать одно из двух заранее определенных значений. Примерами таких переменных могут быть пол пациента, наличие метастаз в лимфоузлах и т.д.
- Порядковые, значения которых можно определенным образом упорядочить и, следовательно, с каждым из них можно сопоставить некоторое целое число – порядковый номер этого значения. Например, проявление болевого синдрома у пациента может выражаться в порядковой шкале – слабая боль, средняя боль, острая боль, невыносимая боль.
- Интервальные, значения которых можно не только упорядочить, но и выразить численно. Например, температура, измеренная в градусах Фаренгейта или Цельсия, образует интервальную шкалу.

Требуется на основании набора показателей $X = (X_1, \dots, X_n)$ составить набор взаимно-независимых показателей $X^* = (X_1, \dots, X_k)$, $k \leq n$, каждый из которых значимо влияет на значения свойств $Y = (Y_1, \dots, Y_m)$.

Методика отбора значимых показателей

Схема решения поставленной задачи заключается в следующем:

1. Необходимо определить, существует ли зависимость между каждым свойством $Y_j, j = \overline{1, m}$ и каждым показателем $X_i, i = \overline{1, n}$. Если свойства $Y_j, j = \overline{1, m}$ не зависят от некоторого показателя X_i , то этот показатель необходимо отбросить из рассмотрения.
2. После этого, необходимо определить, нет ли в получившемся наборе взаимно-независимых показателей. Если такие показатели есть, например, показатели $X_{i_1}, \dots, X_{i_q}$ являются взаимно-независимыми, то необходимо отбросить из рассмотрения все эти показатели, за исключением того, от которого наиболее сильно зависят значения свойств $Y_j, j = \overline{1, m}$.

РАЗДЕЛ IV. ИЗМЕРЕНИЕ, МОДЕЛИРОВАНИЕ И УПРАВЛЕНИЕ В ЭКОЛОГИИ, НАУКАХ О ЧЕЛОВЕКЕ И ОБЩЕСТВЕ

Самыми распространенными методами выявления зависимости между показателями и свойствами являются методы корреляционного анализа и методы экспертной оценки. Однако, будучи примененными по отдельности, они обладают существенными недостатками: большинство методов корреляционного анализа исследует линейную зависимость, что не позволяет говорить об отсутствии связи при нулевом коэффициенте корреляции. При экспертной оценке показателей пациента существенную роль может сыграть субъективное мнение эксперта. Поэтому предлагаемая методика отбора значимых показателей пациента включает корреляционный анализ на начальном этапе анализа и экспертную оценку для окончательного отбора. Такое комбинирование подходов позволяет, с одной стороны, избежать слабостей корреляционного анализа, а именно проблем с нелинейными зависимостями и «ложной» корреляцией. С другой стороны, имея данные корреляционного анализа в качестве отправной точки в своих рассуждениях, эксперты смогут взглянуть на показатели пациентов под новым углом и, при необходимости, скорректировать сложившиеся убеждения.

Предлагаемая методика состоит из следующих шагов:

Рассчитываются коэффициенты корреляции r_{ij} между каждым из показателей

$X_i, i = \overline{1, n}$ и каждым свойством

$Y_j, j = \overline{1, m}$. Соответствующий коэффициент корреляции определяется в соответствии с таблицей 1. Алгоритмы расчета коэффициентов могут быть найдены в [3].

Таблица 1

Типы шкал	Коэффициент корреляции
дихотомическая–дихотомическая	коэффициент четырех-клеточной сопряженности Пирсона
дихотомическая–порядковая	коэффициент рангово-бисериальной корреляции
дихотомическая–интервальная	коэффициент бисериальной корреляции
порядковая–порядковая	коэффициент ранговой корреляции Спирмена
порядковая–интервальная	коэффициент ранговой корреляции Спирмена
интервальная–интервальная	коэффициент корреляции Пирсона

После этого, для каждого показателя $X_i, i = \overline{1, n}$ вычисляется суммарная оценка

$$r_i^S = \sum_{j=1}^m r_{ij}.$$

3. Полученные на предыдущем шаге данные сводятся в следующую таблицу:

4. Таблица 2

Показатель	Y_1	...	Y_k
X_1	a_{11}	...	a_{1k}
...	...	...	...
X_n	a_{n1}	...	a_{nk}

где a_{ij} - интерпретация силы связи, определенная по таблице 3.

Таблица 3

Значение коэффициента корреляции	Сила связи показателей
$r < 0,2$	очень слабая
$0,2 \leq r < 0,3$	слабая
$0,3 \leq r < 0,5$	умеренная
$0,5 \leq r < 0,7$	средняя
$0,7 \leq r$	сильная

5. Полученная таблица (табл. 2) представляется каждому из m экспертов. Эксперт l , основываясь на собственном опыте и с учетом данных таблицы, должен оценить значимость показателей пациента для прогнозирования времени выживания, определив таким образом ранги R_{il} каждого показателя

X_i . Те показатели, которые эксперт считает не влияющими на время выживания, получают ранг 0, а остальные ранжируются последовательными целыми числами, начиная с 1. При этом, больший ранг означает большую значимость показателя.

6. После заполнения таблиц всеми экспертами проводится анализ согласованности экспертных мнений, для чего вычисляется коэффициент конкордации по следующей формуле:

$$W = \frac{12 * S}{m^2 (n^3 - n)},$$

$$S = \sum_{i=1}^n (S_i - S_{cp})^2, \quad S_i = \sum_{j=1}^m r_{ij},$$

где

ВЫБОР ЗНАЧИМЫХ ДЛЯ ПРОГНОЗИРОВАНИЯ ВРЕМЕНИ ВЫЖИВАНИЯ ПОКАЗАТЕЛЕЙ ПАЦИЕНТА

$$S_{cp} = \frac{1}{n} \sum_{i=1}^n S_i$$

7. Коэффициент конкордации W служит мерой общности суждений m ($m > 2$) экспертов. Диапазон изменения коэффициента W от 0 до 1. Если все последовательности рангов совпадают, то $W = 1$. Если все последовательности рангов полностью рассогласованы, то $W = 0$. Увеличение значения коэффициента W означает рост согласованности рангов, а значит, и мнений экспертов [4].
8. Так как эксперты независимы друг от друга в своих суждениях, то и равновероятна любая система рангов, а следовательно, значение коэффициента W является случайной величиной. Для принятия решения о существенности W используется критерий χ_v^2 , где число степеней свободы $v = n - 1$. Для этого вычисляется статистика:

$$\chi_{v \text{ факт}}^2 = \frac{S}{\frac{1}{12} mn(n+1)}$$

- 9.
10. Для заданного уровня значимости α (например, $\alpha = 0,05$) по таблице распределения χ^2 определяется статистика $\chi_{v \text{ табл}}^2$. Если $\chi_{v \text{ табл}}^2 \leq \chi_{v \text{ факт}}^2$, то W существенен при заданном уровне значимости α . Если же это условие не выполняется, то необходимо привлечь большее количество экспертов и вернуться к пункту 3.
11. Если мнения экспертов согласованы, то на основе проведенного ранжирования производится отбор показателей пациента. Те показатели, которые были ранжированы «нулем» большинством экспертов, отбрасываются, а остальные считаются значимо влияющими на время выживания.
12. Определяется взаимосвязь показателей $X_1, \dots, X_n$. Для этого, между всеми оставшимися после предыдущего шага показателями попарно вычисляются коэффициенты корреляции в соответствии с типами их шкал (таблица 1). После расчета значений коэффициен-

тов корреляции определяется сила связи показателей согласно таблице 3. Если сила связи трактуется как средняя или сильная, то показатели считаются связанными.

13. Для каждого показателя $X_i, i = \overline{1, n}$ вычисляется рекомендация c_i , которая может принимать значение «отбросить» либо «оставить» в соответствии со следующим правилом: c_i принимает значение «оставить», если сумма коэффициентов корреляции r_i^S , определенная в пункте 1, для соответствующего показателя $X_i, i = \overline{1, n}$ оказывается больше всех сумм коэффициентов корреляции связанных с ним показателей, и значение «отбросить» в противном случае. Таким образом, на основе данных корреляционного анализа предлагается «оставить» только те показатели пациента, которые наиболее сильно связаны со свойствами, характеризующими время выживания.
14. Полученные данные сводятся в таблицу вида:

Таблица 4

Показатель	Рекомендация	Причина рекомендации
X_1	c_1	
...		
X_n	c_n	

Если некоторому показателю $X_i, i = \overline{1, n}$ соответствует рекомендация «отбросить», то в графе «Причина рекомендации» указывается, что сумма коэффициентов корреляции r_j^S некоторого связанного с ним показателя X_j оказалась больше, чем у данного показателя.

15. Для избежания ошибок, связанных с «ложной» корреляцией, полученная на предыдущем этапе таблица предоставляется каждому из m экспертов. Основываясь на своих знаниях о связи показателей и их физической сущности, а также данных корреляционного анализа, эксперт должен подтвердить или опровергнуть полученную в пункте 10 рекомендацию.

РАЗДЕЛ IV. ИЗМЕРЕНИЕ, МОДЕЛИРОВАНИЕ И УПРАВЛЕНИЕ В ЭКОЛОГИИ, НАУКАХ О ЧЕЛОВЕКЕ И ОБЩЕСТВЕ

16. После получения мнений всех экспертов производится окончательный отбор показателей – те показатели, которые были «отброшены» более чем половиной экспертов исключаются из рассмотрения, а оставшиеся показатели считаются значимо влияющими на время выживания пациента.

Программный комплекс

Представленная методика была реализована в соответствующем интернет - приложении, разработанном с использованием связки Apache+MySQL+PHP. Разработанное приложение позволяет загружать статистические данные в формате XML или CSV, рассчитывать корреляционные связи между показателями и свойствами, проводить экспертную оценку, оценку согласованности мнений экспертов и, при необходимости, привлекать новых экспертов, формировать отчет по отобранным показателям.

СПИСОК ЛИТЕРАТУРЫ

1. Боженко В.К., Сотников В.М. Использование многопараметрических методов анализа информации в онкологии. // Вестник Российского Научного Центра Рентгенодиагностики, вып. 6., Москва, 2006.
2. Lee, Elisa T. and Wang, John. Statistical Methods For Survival Data Analysis. 3 edition. s.l. // Wiley-Interscience, 2003.
3. Т.А. Баранова. Многомерные статистические методы. Корреляционный анализ. // Издательство ИГХТУ, Иваново, 2007.
4. Емалетдинова Л.Ю., Габитов Р.И. Автоматизация экспертного оценивания деятельности специалистов лечебного учреждения. // Вестник КГТУ им. А.Н.Туполева, Казань, 1997.

Аспирант **Д.Ю. Стрункин** тел. 8-904-666-99-02, dstrunkin@yandex.ru - каф. Прикладной Математики и Информатики КГТУ им. А.Н. Туполева.

УДК 528.85

РЕЗУЛЬТАТЫ ИСПОЛЬЗОВАНИЯ КОРОТКОИМПУЛЬСНОЙ СИСТЕМЫ КОНТРОЛЯ ЛЕСНОЙ СРЕДЫ

О.Н. Очиров, Б.Ч. Доржиев, А.В. Базаров

Проведена оценка коэффициента ослабления леса на частоте 10 ГГц с использованием короткоимпульсной системы контроля лесной среды. Предложен метод построения радиотомограмм тестового участка леса при однопозиционном сканировании наносекундным радаром.

Ключевые слова: система контроля, лесная среда.

Введение

Сверхкороткие импульсы находят широкое применение в таких областях радиоэлектроники как сверхширокополосная радиосвязь, радиолокация, системы точного позиционирования и т.д. [1]. Вместе с тем несомненный интерес представляет их использование в системах дистанционного зондирования, в частности, при исследовании лесных сред методами сверхкороткоимпульсной радиолокации. Сверхкороткоимпульсная радиолокация (СКИРЛ) является одной из ветвей технологии сверхширокополосных (СШП) систем. Наиболее характерными специфическими особенностями СКИРЛ [2] являются высокая разрешающая способность по дальности, определяемая шириной используемого частотного спектра и эффект контрастности отражений от объектов. Указанные особенности приводят к другому, чем в традиционной локации, характеру отраженных сигналов от

местности, создавая рельефную картину из отдельных импульсов отраженных от местных предметов и характерных участков местности, в том числе и на предельно малых дальностях.

Целью настоящей работы является экспериментальное выявление особенностей СКИРЛ лесных сред, связанных, в основном, с их высокой разрешающей способностью, а также оценка их ослабляющих свойств.

Оценка коэффициента ослабления леса на частоте 10 ГГц. Как указывается во многих работах, например [3], лес считается наиболее сложным объектом для построения моделей рассеяния волн. Основой для построения таких моделей являются результаты экспериментальных исследований, полученных методами традиционной радиолокации. Новые результаты, полученные при короткоимпульсной локации лесных сред, послужат для дальнейшего развития моделей и их практи-