

В.И. Чернов, Н.А. Гладских, А.В. Чернов, С.И. Штаньков
МЕТОДИКА ФОРМИРОВАНИЯ
ДИАГНОСТИЧЕСКИ ЗНАЧИМЫХ ПРИЗНАКОВ ПРИ РАСЧЕТЕ
ВЕРОЯТНОСТИ ПОВТОРНОГО ИНСУЛЬТА

ГОУ ВПО ВГМА им. Н.Н.Бурденко Минздравсоцразвития России

Резюме. При расчете вероятности повторного инсульта необходимо получить перечень наиболее информативных признаков. При разработке словаря признаков приходится сталкиваться с некоторыми ограничениями. Одно из них состоит в том, что в словарь могут быть включены только те признаки, для которых имеется априорная информация, достаточная для описания классов на языке этих признаков. Другое ограничение заключается в том, что некоторые из признаков нецелесообразно включать в априорный словарь ввиду того, что они малоинформативны.

Ключевые слова: словарь диагностически значимых признаков, методика Кульбака, расчет вероятности повторного инсульта.

Актуальность. Проблема формирования словаря диагностически значимых признаков является актуальной в настоящее время и применение данных методик значительно снижает вероятность повторного инсульта [1, 2, 3, 4, 5].

Проблема цереброваскулярных заболеваний и инсульта в последние годы становится все более актуальной. Ежегодно в мире мозговой удар настигает более 15 млн. человек. В России каждый год регистрируется более 500 тысяч случаев острых нарушений мозгового кровообращения (ОНМК).

Во всех странах инсульт – лидирующая причина первичной инвалидизации: более половины пациентов, перенесших мозговую катастрофу, нуждаются в той или иной степени ухода за собой. Вторичная профилактика – гораздо более специализированная область лечения инсультов. Многочисленные клинические испытания подтвердили эффективность мероприятий стратегии высокого риска в профилактике повторных инсультов.

Материал и методы исследования. На предварительном этапе при расчете вероятности повторного инсульта необходимо получить перечень наиболее информативных признаков. Построенный словарь признаков должен явиться информативной базой для расчета вероятности повторного инсульта. Определение словаря признаков возможно с использованием перечисленных ниже подходов.

1. Игровой подход к построению словаря признаков.
2. Метод, основанный на сравнении апостериорных вероятностей.
3. Метод, основанный на сравнении вероятностных характеристик признаков.
4. Метод, основанный на определении количества информации.
5. Метод, базирующийся на определении информативности Кульбака.

Методика расчета информативности признаков по Кульбаку базируется на определении диагностических коэффициентов, рассчитанных для основной и контрольной групп пациентов.

Диагностический коэффициент представляется в виде логарифма отношения вероятностей проявления данного признака в основной и контрольной группе ($p(X_{ij}|A_1)$ и $p(X_{ij}|A_2)$ соответственно) и умноженный на 100:

$$ДК = 100 \lg \frac{P(x_j / A_1)}{P(x_j / A_2)}$$

Диагностические коэффициенты представляют собой чаще всего двузначные или однозначные положительные или отрицательные числа. Положительными они являются в случае преобладания вероятности $p(X_{ij}|A_1)$, находящейся в числителе, отрицательными — в случае преобладания вероятности $p(X_{ij}|A_2)$. Чем больше величина диагностического коэффициента, тем больше дифференциально-диагностической информации, т. е. информации о преобладании вероятности одного из диагнозов, он несет. Однако информативность каждого значения признака зависит также от частоты, с какой встречается это значение при каждом из заболеваний, т. е. от величин $p(X_{ij}|A_1)$ и $p(X_{ij}|A_2)$.

Алгоритм формирования словаря информативных признаков состоит из следующих этапов:

1. Формирование основной и контрольной группы пациентов;
2. Расчет вероятностей проявления признака в основной и контрольной группах $p(X_{ij}|A_1)$ и $p(X_{ij}|A_2)$;
3. Расчет диагностических коэффициентов для признаков $ДК()$;
4. Вычисление информативности для заданного значения признака $ДК() [p(X_{ij}|A_1) - p(X_{ij}|A_2)]$;
5. Вычисление информативности признака $I(x_i) = ДК() [p(X_{ij}|A_1) - p(X_{ij}|A_2)]$;
6. Отбор признаков, имеющих наибольшее значение $I(x_i)$.

Полученные результаты и их обсуждение. Пользуясь предложенной методикой, на основе сформированной базы данных были рассчитаны диагностические коэффициенты и значения информативности по каждому из признаков:

X1 – нарушение сознания; X2 – гемианопсия; X3 – парез в руке; X4 – парез в ноге; X5 – расстройство чувствительности (гемигипостезия); X6 – симптом отрицания (анозогнозия); X7 – афазия; X8 – нарушение ритма сердца; X9 – сахарный диабет; X10 – показатели глюкозы крови на момент инсульта; X11 – ультразвуковая доплерография (УЗДГ); X12 – возраст; X13 – пол; X14 – АД; X15 – холестерин; X16 – ИБС; X17 – Локализация очага по бассейнам; X18 – частота подтипов; X19 – Тяжесть инсульта по Ренкину; X20 - Баллы по Бартелу.

Для определения вероятностных оценок рецидива инсульта целесообразно использовать формулу Байеса, которую иногда называют теоремой об обратной вероятности или теоремой гипотез.

Формула Байеса имеет следующий вид:

$$P(A_1 / x_1) = \frac{P(A_1)P(x_1 / A_1)}{\sum_k P(A_k)P(x_1 / A_k)}$$

По существу задача диагностики состоит в том, чтобы установить диагноз, используя тот минимум доступной диагностической информации, который достаточен для достижения необходимой надежности диагноза. Это обычно требует использования не одного симптома, а набора симптомов (симптомокомплекса). Такой подход может быть назван «многомерным» подходом к установлению диагноза, так как при нем одновременно используют много признаков.

Таким образом, методика расчета вероятности повторного инсульта состоит из следующих этапов:

1. Формирование пространства признаков, которые позволяют полностью идентифицировать состояние объекта моделирования.
2. Формирование словаря информативных признаков на основе критерия Кульбака.
3. Расчет вероятности повторного инсульта на основе формулы Байеса.
4. Формирование рекомендаций по дальнейшему лечению и профилактике.

Под нашим наблюдением находился 191 больной, перенесший один и более ишемический инсульт. За 38 пациентами осуществлялось ретроспективное наблюдение, 153 пациента наблюдались с первого момента развития заболевания. Контрольную группу составили 80 больных, перенесших один инсульт, 111 больных с повторными нарушениями мозгового кровообращения – основную группу. Их обследование и лечение проходило на базе неврологического отделения для больных с нарушениями мозгового кровообращения Воронежской областной клинической больницы №1, с 2000 года по 2007 год.

Для анализа возраста пациенты основной и контрольной группы были разбиты на вариационные ряды, где шаг интервала и их количество определялись по формулам:

$$k \geq 1 + 3.32 \lg N, \quad (1)$$

где k – количество интервалов, N – количество пациентов.

$$h \geq (n_x - n_1) / k, \quad (2)$$

где h – шаг интервала, n_x и n_1 , крайние значения.

Литература:

1. Проблема оптимизации численности врачебных кадров на региональном уровне и возможность ее решения посредством создания интерактивной системы сбора информации о трудоустройстве выпускников вуза / И.Я. Львович, Н.А. Гладских, С.Н. Семенов, Е.Б. Смолькин // Прикладные информационные аспекты медицины – 2006. – Т.9. – № 2. Режим доступа: <http://www.vsm.a.ac.ru/publ/vestnik/priam/009-2/Site/index.html#5>.
2. Гладских Н.А. Алгоритм методики прогнозирования медико-демографических показателей на основе методов экспоненциального сглаживания и множественной регрессии // Научно-медицинский вестник Центрального Черноземья. – 2007. - №7. Режим доступа: <http://www.vsm.a.ac.ru/publ/vestnik/vest/027/Site/index.html#19>.

3. Львович И.Я., Гладских Н.А., Семенов С.Н. Классификация объектов прогнозирования демографических показателей населения региона методами кластерного анализа // Прикладные информационные аспекты медицины. – 2006. – Т.9. – № 2. Режим доступа: [http://www.visma.ac.ru/publ/ vestnik/priam/009-2/Site/index. html#19](http://www.visma.ac.ru/publ/vestnik/priam/009-2/Site/index.html#19).

4. Программная реализация методики прогнозирования медико-демографических данных [Электронный ресурс] / И.Я. Львович, С.Н. Семенов, Н.А. Гладских, С.С. Пронин // Научно-медицинский вестник Центрального Черноземья.-2007.-№7. Режим доступа: <http://www.visma.ac.ru/publ/vestnik/vest/027/Site/index.html#19>.

5. Методика прогнозирования медико-демографических показателей на базе методов экспоненциального сглаживания и множественной регрессии / И.Я. Львович, С.Н. Семенов, Н.А. Гладских, С.С. Пронин // Научно-медицинский вестник Центрального Черноземья. – 2007. - №7. Режим доступа: <http://www.visma.ac.ru/publ/vestnik/vest/027/Site/index.html#20>.