Модели-классификаторы в медицине

Возжаева И.В.
Вычислительный Центр ДВО РАН (Хабаровск)

В наше время математическое моделирование – компонент, активно участвующий почти во всех отраслях науки, в том числе и в медицине.

При изучении здоровья населения возникает задача поиска методов, опирающихся на изучаемые факторы, предположительно имеющие влияние на результаты медицинских обследований, а также задача нахождения оптимального инструмента, с помощью которого будет реализован этот метод и создана модель. Это необходимо не только для установления диагноза, но и, что не менее важно, выяснения закономерностей между перечнем результатов обследований пациента и выявлением наличия болезни (т.е. входными и выходными данными). Или, проще говоря, на практике с помощью математического моделирования можно определить, влияет ли тот или иной факт на итог.

Учитывая, что при таких заболеваниях, как рассеянный склероз (РС) обследование пациента требует далеко не малых денежных затрат, то создание подобных моделей является актуальной задачей современности.
Не стоит забывать и о том, что тот же РС – болезнь до конца не изученная, на сегодняшний день не излечимая (возможно только предотвращение рецидива) и нахождение факторов, влияющих на скорость прогрессирования заболевания - задача не просто интересная науке, а жизненно важная для людей, страдающих этим недугом.

Таким образом, перед нами была поставлена цель: найти метод, способный классифицировать по трем категориям скорость развития РС (медленная, средняя, быстрая) и, применив его на практике, выявить зависимость между конечным результатом и предоставленными для исследовательской работы данными обследований пациентов.

Метод, используемый ранее для решения выше описанной задачи, требовал проведения расчетов с участием многих величин, однако метод, описанный нами, в процессе практического применения позволил сократить их число почти вдвое, за ненадобностью.

Реализация метода была осуществлена с помощью построения модели обучающейся нейросети. Этот выбор обоснован тем, что РС – заболевание головного мозга, а нейронные сети наиболее адекватно отражают его деятельность. Также был учтен тот факт, что исследуемое заболевание не имеет четко определенных закономерностей развития.

Предложенная база данных содержит следующие результаты исследований пациентов с РС:

1. Номер, присвоенный пациенту в базе данных;
2. Пол;
3. Оценка восьми функциональных систем от 0 – норма, до 6 – полная или почти полная дисфункция системы (кроме функции мышления – оценка по шкале от 0 до 5):
4. Сумма неврологического дефицита (суммарная оценка всех систем);
5. Сумма пораженных систем (количество пораженных систем);
6. Оценка инвалидности (EDSS) по шкале Куртцке;
7. Продолжительность болезни (в годах);
8. Возраст пациента в начале РС;
9. Начальные симптомы по восьми функциям организма;
10. Коэффициент экзацербации, характеризующий количество обострений за год;
11. Скорость прогрессирования РС в баллах;
12. Скорость (темп) прогрессирования РС (по трем категориям: 1 – медленный, 2 – умеренный, 3 – быстрый).

Результаты обследований, отраженные в базе данных уже были выведены с помощью статистических анализа и оценки.

Входными данными являются пункты №2 – 10 . Темп прогрессирования РС – выходные данные.

Специфика предоставленных данных задала способ построения и алгоритм обучения нейронной сети. Построение установлено с обратной связью, а обучение осуществляется с «учителем» (т. е. от нейронной сети требуется найти закономерности между входными и выходными данными; т. о. входные данные представлены в виде вектора, скалярным значением которого является конечный результат - пункт №12) на основе алгоритма обратного распространения, который заключается в распространении сигналов ошибки от выходов нейросети к ее входам, в направлении, обратном прямому распространению сигналов в обычном режиме работы.

Преимущество такой схемы в минимизации нейронной сети из-за многократного участия каждого нейрона в обработке данных, что облегчает процесс обучения.

При этом ошибка сети зависит от ее конфигурации – совокупности всех ее синаптических весов. Но эта зависимость не прямая, а опосредованная.

Иными словами, в общем виде функция ошибки имеет вид:

$$E(w) = E\{x^\alpha, y^\alpha, y(x^\alpha, w)\}.$$

Здесь $\{x^\alpha, y^\alpha\}$ – набор примеров (т.е. пар входов-выходов), на которых обучается нейросеть, а $\{y(x^\alpha, w)\}$ – реальные значения выходов нейросети, зависящие от конкретных значений ее синаптических весов.

Когда функционал ошибки задан, и задача сводится к его минимизации, можно предложить, например, следующую итерационную процедуру подбора весов: $w^{\tau+1} = w^\tau - \eta^\tau \frac{\partial E}{\partial w}$ или, что то же самое: $w_{ij}^{\tau+1} = w_{ij}^\tau - \eta^\tau \frac{\partial E}{\partial w_{ij}}$.

Здесь $\eta^{\tau} << \|w\|$ – темп обучения на шаге τ. Можно показать, что постепенно уменьшая темп обучения описанная выше процедура приводит к нахождению локального минимума ошибки.

Исторически наибольшую трудность на пути к эффективному правилу обучения многослойных персептронов вызвала процедура эффективного расчета градиента функции ошибки $\frac{\partial E}{\partial w}$. Дело в том, что ошибка сети определяется по ее выходам, т.е. непосредственно связана лишь с выходным слоем весов. Вопрос состоял в том, как определить ошибку для нейронов на скрытых слоях, чтобы найти производные по соответствующим весам. Нужна была процедура передачи ошибки с выходного слоя к предшествующим слоям сети, в направлении обратном обработке входной информации. Поэтому такой метод, когда он был найден, получил название метода обратного распространения ошибки.

Между тем, в случае дифференцируемых функций активации рецепт нахождения производных по любому весу сети дается так называемым цепным правилом дифференцирования. Суть метода back-propagation в эффективном воплощении этого правила.

Разберем этот метод подробнее. Обозначим входы n-го слоя нейронов $x_j^{[n]}$. Нейроны этого слоя вычисляют соответствующие линейные комбинации:

$$a_i^{[n]} = \sum_j w_{ij}^{[n]} x_j^{[n]}$$

и передают их на следующий слой, пропуская через нелинейную функцию активации (для простоты – одну и ту же, хотя это необязательно):

$$x_i^{[n+1]} = f\left(a_i^{[n]}\right).$$

Для построения алгоритма обучения нам надо знать производную ошибки по каждому из весов сети:

$$\frac{\partial E}{\partial w_{ij}^{[n]}} = \frac{\partial E}{\partial a_i^{[n]}} \frac{\partial a_i^{[n]}}{\partial w_{ij}^{[n]}} \equiv \delta_i^{[n]} x_j^{[n]}.$$

Таким образом вклад в общую ошибку каждого веса вычисляется локально, простым умножением невязки нейрона $\delta_i^{[n]}$ на значение соответствующего входа.

Входы каждого слоя вычисляются последовательно от первого слоя к последнему во время прямого распространения сигнала:

$$x_i^{[n+1]} = f\left(\sum_j w_{ij}^{[n]} x_j^{[n]}\right),$$

а невязки каждого слоя вычисляются во время обратного распространения ошибки от последнего слоя (где они определяются по выходам сети) к первому:

$$\delta_i^{[n]} = f'\left(a_i^{[n]}\right)\sum_k w_{ki}^{[n+1]} \delta_k^{[n+1]}.$$

Последняя формула получена применением цепного правила к производной

$$\frac{\partial E}{\partial a_i^{[n]}} = \sum_k \frac{\partial E}{\partial a_k^{[n+1]}} \frac{\partial a_k^{[n+1]}}{\partial x_i^{[n+1]}} \frac{\partial x_i^{[n+1]}}{\partial a_i^{[n]}}$$

и означает, что чем сильнее учитывается активация данного нейрона на следующем слое, тем больше его ответственность за общую ошибку.

Рассмотрим процедуру обратного распространения. Минимизируемой целевой функцией ошибки нейросети является величина:

$$E(w) = \frac{1}{2}\sum_{j,p} (y_{j,p}^{(N)} - d_{j,p})^2 \quad (1)$$

где $y_{j,p}^{(N)}$ – реальное выходное состояние нейрона j выходного слоя N нейронной сети при подаче на ее входы p-го образа; d_{jp} – идеальное (желаемое) выходное состояние этого нейрона.

Суммирование ведется по всем нейронам выходного слоя и по всем обрабатываемым сетью образам. Минимизация ведется методом градиентного спуска, что означает подстройку весовых коэффициентов следующим образом:

$$\Delta w_{ij}^{(n)} = -\eta \cdot \frac{\partial E}{\partial w_{ij}} \qquad (2)$$

Здесь w_{ij} – весовой коэффициент синаптической связи, соединяющей i-ый нейрон слоя $n-1$ с j-ым нейроном слоя n, η – коэффициент скорости обучения, $0<\eta<1$.

Обобщим (1) и (2):

$$\frac{\partial E}{\partial w_{ij}} = \frac{\partial E}{\partial y_j} \cdot \frac{dy_j}{ds_j} \cdot \frac{\partial s_j}{\partial w_{ij}} \qquad (3)$$

Здесь под y_j, как и раньше, подразумевается выход нейрона j, а под s_j – взвешенная сумма его входных сигналов, то есть аргумент активационной функции. Так как множитель dy_j/ds_j является производной этой функции по ее аргументу, из этого следует, что производная активационной функция должна быть определена на всей оси абсцисс.

Третий множитель $\partial s_j/\partial w_{ij}$ равен выходу нейрона предыдущего слоя $y_i^{(n-1)}$.

Что касается первого множителя в (3), он легко раскладывается следующим образом:

$$\frac{\partial E}{\partial y_j} = \sum_k \frac{\partial E}{\partial y_k} \cdot \frac{dy_k}{ds_k} \cdot \frac{\partial s_k}{\partial y_j} = \sum_k \frac{\partial E}{\partial y_k} \cdot \frac{dy_k}{ds_k} \cdot w_{jk}^{(n+1)} \qquad (5)$$

Здесь суммирование по k выполняется среди нейронов слоя $n+1$.

Введя новую переменную

$$\delta_j^{(n)} = \frac{\partial E}{\partial y_j} \cdot \frac{dy_j}{ds_j} \qquad (6)$$

мы получим рекурсивную формулу для расчетов величин $\delta_j^{(n)}$ слоя n из величин $\delta_k^{(n+1)}$ более старшего слоя $n+1$.

$$\delta_j^{(n)} = \left[\sum_k \delta_k^{(n+1)} \cdot w_{jk}^{(n+1)} \right] \cdot \frac{dy_j}{ds_j} \qquad (7)$$

Для выходного же слоя

$$\delta_l^{(N)} = (y_l^{(N)} - d_l) \cdot \frac{dy_l}{ds_l} \qquad (8)$$

Теперь мы можем записать (2) в раскрытом виде:

$$\Delta w_{ij}^{(n)} = -\eta \cdot \delta_j^{(n)} \cdot y_i^{(n-1)} \qquad (9)$$

Иногда для придания процессу коррекции весов некоторой инерционности, сглаживающей резкие скачки при перемещении по поверхности целевой функции, (9) дополняется значением изменения веса на предыдущей итерации

$$\Delta w_{ij}^{(n)}(t) = -\eta \cdot (\mu \cdot \Delta w_{ij}^{(n)}(t-1) + (1-\mu) \cdot \delta_j^{(n)} \cdot y_i^{(n-1)}) \qquad (10)$$

где μ – коэффициент инерционности, t – номер текущей итерации.

Таким образом, полный алгоритм обучения НС с помощью процедуры обратного распространения строится так:

1. Подать на входы сети один из возможных образов и в режиме обычного функционирования НС, когда сигналы распространяются от входов к выходам, рассчитать значения последних. Напомним, что

$$s_j^{(n)} = \sum_{i=0}^{M} y_i^{(n-1)} \cdot w_{ij}^{(n)} \qquad (11)$$

где M – число нейронов в слое n-1 с учетом нейрона с постоянным выходным состоянием $+1$, задающего смещение; $y_i^{(n-1)}=x_{ij}^{(n)}$ – i-ый вход нейрона j слоя n.

$y_j^{(n)} = f(s_j^{(n)})$, где $f(...)$ – сигмоид (12)

$y_q^{(0)}=I_q$, (13)

где I_q – q-ая компонента вектора входного образа.

2. Рассчитать $\delta^{(N)}$ для выходного слоя по формуле (8).

Рассчитать по формуле (9) или (10) изменения весов $\Delta w^{(N)}$ слоя N.

3. Рассчитать по формулам (7) и (9) (или (7) и (10)) соответственно $\delta^{(n)}$ и $\Delta w^{(n)}$ для всех остальных слоев, $n=N-1,...1$.

4. Скорректировать все веса в НС

$$w_{ij}^{(n)}(t) = w_{ij}^{(n)}(t-1) + \Delta w_{ij}^{(n)}(t) \qquad (14)$$

5. Если ошибка сети существенна, перейти на шаг 1. В противном случае – конец. [1]

По итогам проведенных экспериментов оказалось, что на результат не оказывают влияния возраст пациента и возраст в начале болезни. Также было выяснено, что половая принадлежность значительно влияет на итог анализа.

Предложенный метод показал, что в разрез с общепринятым мнением медков скорость прогрессирования РС можно определить без участия начальных симптомов болезни, зато на результат влияет не только сумма неврологического дефицита, но и оценка каждой из восьми функциональных систем в отдельности.

При применении этого метода на практике подтвердилась его оптимальность для решения поставленной задачи. Единственным недостатком предложенного метода является необходимость в более обширной базе данных, так как чем больше примеров задано (т.е. известных пар «вход-выход»), тем больше результативность нейросети.

Литература

1. Нейрокомпьютинг и его применения в экономике и бизнесе – А.А.Ежов, С.А.Шумский. Москва,1998 г.