

О ПРИМЕНЕНИИ НЕЙРОСЕТЕВОГО МЕТОДА СИНТЕЗА ДИСКРИМИНАНТНЫХ ФУНКЦИЙ ДЛЯ ПРОГНОЗИРОВАНИЯ ИСХОДА НЕРВНО-ПСИХИЧЕСКИХ ЗАБОЛЕВАНИЙ У ЛИЦ МОЛОДОГО ВОЗРАСТА

Рассматривается одна из медицинских задач прогнозирования методами синтеза дискриминантных функций на примере группы призывников в вооруженные силы и увольнения в запас.

There is being considered one of the medicine problems of forecasting by the methods of synthesis of the discriminant functions on the example of the group of conscripts into the armed forces and transfer to the reserve.

Для прогнозирования исхода течения заболевания применяются классические методы математической статистики, к которым относятся методы факторного анализа, линейной и нелинейной регрессии [1 – 4]. Эффективные прогнозные оценки в этих методах достигаются при выполнении условий гауссовой распределения, однородности и независимости случаев, статистическую проверку которых можно осуществить на достаточно представительной выборке, включающей не менее чем 60 наблюдений. Однако в результате проверки может быть выяснено, что устойчивой функции распределения не существует или же ее аппроксимация гауссовой кривой невозможна [5]. В тех редких случаях, когда эти статистические условия выполняются без привлечения особой процедуры подгонки данных, на основе которых затем находится прогнозная модель, то содержание этой модели, выраженное на математическом языке, не соответствует клиническому мышлению врача.

По мнению Неймарка Ю. И. и соавторов (1991), указанные недостатки классических методов могут быть устранены, если прогнозные модели искать в классе логических функций, используя для этого минимальные дизъюнктивные формы логических высказываний. Однако для сложных многофакторных прогнозных моделей этот метод требует больших вычислительных затрат, снижение которых возможно при помощи нейросетевых методов синтеза [5].

Для нейросетевых методов требуется составить обучающую выборку из тех случаев, классификация которых не вызывает сомнений и разногласий экспертов. Случаи должны быть представлены результатами клинических и лабораторных исследований, осуществляемых в соответствии с принятой технологией. Результаты исследований, включаемые в выборку, не должны содержать пропущенных значений, однако в исключительных случаях, когда статистика мала, пропуски допускаются, хотя достоверность таких прогнозных моделей будет снижаться.

Проведенные исследования показали, что нейросетевые методы способны распознавать прогнозные факторы уже при 5 классификационных случаях для каждого из двух течений заболевания без осложнений и с осложнениями; чем больше количество случаев, тем больше достоверность прогнозной модели.

В качестве примера рассмотрим исследование группы юношей призывающихся на службу в Вооруженные силы АО Пензенской области и уволенных в запас с 1994 по 2009 года.

Для классификации течения заболеваний была подготовлена обучающая выборка, состоящая из 8 случаев без осложнений, кодируемых 1, и 5 случаев осложнений, кодируемых 0. Эти случаи представлены данными клинико-лабораторных исследований, проведенных в течении 10 дней. Их структура приведена в табл. 1.

Нейросетевой метод построения моделей прогнозирования течения заболевания включает несколько этапов. На первом из них определяются пороговые величины и тип пороговой функции, при помощи которых показатели преобразуются в логические переменные; тип пороговой функции обозначен 0, когда показатель кодируется 0, если значение переменной превышает порог или равно ему, и 1 – в ином случае. Приведенные в табл. 1 пороговые величины находятся из условия минимума числа ошибок (ЧО), допускаемых на обучающей выборке однофакторной моделью, состоящей из одной логической переменной.

Из этого списка следует удалить показатели, для которых ЧО равно числу случаев одного из двух исходов; такие клинические и лабораторные исследования, как возраст и лейкоциты, травматизм и наличие белка в моче не обладают разделительной силой, поскольку их значения принимают примерно одинаковые значения для всех случаев выборки.

Структура показателей

Таблица 1

N	Наименование показателя	Порог	Функция	ЧО
1.	Возраст	26	1	5
2.	Сопутствующая патология	-	-	-
3.	Длительность заболевания	4	0	4
4.	Стадия заболевания	-	-	-

5.	Объем болезни	-	-	5
6.	Гемоглобин, г/л	90.9	1	4
7.	Эритроциты, 1012	3.3	1	4
8.	Лейкоциты, 109	3.1	1	5
9.	СОЭ, мм/час	18.0	0	4
10.	Остаточный азот, мкмоль/л	21.4	0	3
11.	Сахар, ммоль/л	4.6	0	4
12.	Общий билирубин, мкмоль/л	15.4	0	2
13.	Мочевина, ммоль/л	6.5	0	4
14.	Общий белок, г/л	63.7	1	4
15.	Фибриноген, г/л	1.3	1	4
16.	ПТИ, %	70.6	1	3
17.	Белок в моче	-	-	5
18.	Изменения ЭКГ	-	-	-
19.	Изменения спирограммы	-	-	-

Примечание: ЧО – частота ошибок показателя на обучающей выборке

После этого следует убедиться в том, что в обучающей выборке нет одинаковых случаев, принадлежащих разным исходам. Наличие таких случаев свидетельствует о том, что состав показателей, вошедших в обучающую выборку, не полон и его следует расширить за счет новых переменных, которые по мнению эксперта могут оказывать влияние на исход болезни. Если перечень таких показателей исчерпан, а противоречие не устранено, то в ущерб достоверности прогнозной модели следует удалить один из таких случаев.

Из оставшихся переменных наибольшей разделительной силой обладает показатель общего билирубина, который обеспечивает 2 ошибки классификации.

Для увеличения достоверности прогнозных моделей необходимо усложнить найденные модели так, чтобы включали дополнительные переменные. Из существующих двухфакторных вариантов можно выделить 7 частных моделей, число ошибок которых уже равно 1.

Дальнейшее усложнение приводит к появлению 26 трехфакторных моделей, число ошибок которых равно 0.

Анализируя приведенные комбинации факторов моделей прогнозирования в табл. 2, можно заметить, что в прогнозные модели не вошли такие лабораторные и клинические исследования, как длительность болезни, гемоглобин, эритроциты, мочевина и изменения спирограммы. Наиболее информативными оказывается тройка факторов 10-14-16 (остаточный азот, общий белок, ПТИ), обеспечивающая наибольшее число вариантов моделей. Заметим, что при большем количестве вариантов достигается более высокая достоверность прогноза. Однако такое большое число моделей, равное 10, могло произойти из-за пропущенных значений, оказавшихся в показателе остаточного азота при составлении обучающей выборки. Поэтому для тех случаев, в которых будут присутствовать пропуски этого параметра, следует воспользоваться одной из 16 оставшихся моделей.

Дополнительным критерием отбора вариантов может быть стоимость исследований, в соответствии с которым предпочтение отдается более экономичным моделям прогноза.

Рассмотрим лишь два варианта моделей, один из которых обеспечивается комбинацией переменных 10-14-16, а другой – комбинацией 2-11-12. Как уже отмечалось, предпочтительными оказываются комбинации с наибольшим числом и вариантов моделей. Тогда возникает проблема оценки согласованности решений и выбора критериального уровня, при превышении которого они принимаются. Для решения этой проблемы вводится коэффициент согласованности (КС), равный отношению числа моделей, проголосовавших за принятое решение, к общему их числу; величина КС изменяется от 0,5 в случае неопределенности, до 1 – при полной согласованности.

В табл. 3 приводятся 4 модели по схеме 10-14-16, в которой символ "плюс" означает, что значение переменной превышает или равно соответствующему порогу из табл. 7, а символ "минус" – иное состояние; для всех моделей этой таблицы величина КС=1.

Таблица 2

Комбинация факторов моделей прогнозирования

№	Клинические исследования		Лабораторные исследования								ЧМ
	2	18	9	10	11	12	14	15	16		

1.		$ \begin{array}{ccc} + & + & + \\ + & + & \\ + & + & + \\ + & + & + \\ + & + & \\ + & + & \\ + & + & \\ \end{array} $	2
2.			4
3.			10
4.			6
5.	+		2
6.	+		1
7.	+		1

Примечание: ЧМ – число моделей для данной комбинации факторов

Таблица 3

Интерпретация трехфакторных моделей по схеме 10-14-16

Факторы				
№	10	14	16	ПР
1.	+	+	+	0
2.	-	-	+	0
3.	-	+	-	0
4.	-	+	+	1

Примечание: ПР – принимаемое решение: 0 – осложненное, 1 – нормальное течение

Оставшиеся 4 комбинации факторов не включены в эту таблицу, поскольку их КС меньше заданного порога, величина которого выбрана нами равной 0,9. В том случае, если комбинация факторов 10-14-16 отсутствует в табл. 9 или же значения одного из показателей этой комбинации пропущены, следует воспользоваться схемой 2-11-12, содержащей 7 вариантов факторов, для которых величина КС также равна 1 (табл. 4).

Заметим, что в схеме 2-11-12, обеспечивающей всего 2 модели, существует лишь один несогласованный вариант.

Таблица 4

Интерпретация трехфакторных моделей по схеме 2-11-12

Факторы				
№	2	11	12	ПР
1.	-	+	-	0
2.	-	+	+	0
3.	-	-	-	1
4.	+	+	+	0
5.	+	+	-	1
6.	+	-	+	0
7.	+	-	-	1

Примечание: ПР – принимаемое решение: 0 - осложненное, 1 - нормальное течение

Результаты проверки прогностической таблицы на контрольном примере состоящем из двух случаев с нормальным и осложненным исходом, показатели 2-11-12 которых равны соответственно: (+)-4,3-7,4 и (+)-6,8-13,3, подтверждают работоспособность данного подхода.

Прогностические таблицы удобно реализовать на компьютере, однако ими можно воспользоваться и при его отсутствии. Для повышения достоверности прогнозов эти таблицы можно модифицировать, если будут новые дополнительные данные о классифицированных случаях.

Таким образом, нейросетевой метод обеспечивает безошибочную классификацию обучающей выборки при допустимом числе ошибок на экзаменационной выборке.

Список использованных источников

1. Brown D. // Amer. J. Surg. – 1965. – V. 109. – P. 521 – 525.
2. Сотникова А. А. Системный анализ и прогноз влияния эпидемий гриппа на нервно-психические заболевания на фоне солнечной активности / А. А. Сотникова, В. Г. Камбург, М. И. Гринцов // Информационно-вычислительные технологии и их приложения : сб. статей X Междунар. научной конф. – С. 255 – 259.
3. Сотникова А. А., Камбург В. Г. Модельный прогноз заболевания психопатии на фоне солнечной активности // Вісник Хмельницького національного університету. Технічні науки. – 2009. – № 5. – С. 97 – 100.
4. Камбург В. Г., Сотникова А. А., Бочкарёва О. В. Модельный прогноз эпидемий гриппа и его влияние на нервно-психические заболевания на фоне солнечной активности // Вісник Хмельницького

національного університету. Технічні науки. – 2009. – № 5. – С. 101 – 106.

5. Щетинин В. Г. Исследование методов многорядной самоорганизации разделяющих функций и разработка дискриминаторов : автореф. дис. на соискание научной степени канд. техн. наук / В. Г. Щетинин. – Пенза, 1996. – 18 с.