

Литература

1. Kushwaha, Kss & Kushwaha, Neelottama & Rai, Arti. (2010). Recent Trends on the Future of Graduate Education in the Pharmaceutical Sciences and Research. Journal of young pharmacists: JYP. 2. 206-12. 10.4103/0975-1483.63173.
2. Аладышева, Ж. И., Береговых, В. В., Пятигорская, Н. В., Самылина, И. А. Актуальные вопросы современного фармацевтического образования в Российской Федерации // Фармация. 2013. – № 1. – С. 3-7.

ИТЕРАТИВНЫЙ МЕТОД ОПРЕДЕЛЕНИЯ ОПТИМАЛЬНЫХ МОДЕЛЕЙ МНОЖЕСТВЕННОЙ РЕГРЕССИИ ПРИ ИХ ПРЯМОМ ПЕРЕБОРЕ В РАМКАХ НАУЧНО-ОРИЕНТИРОВАННОГО И УГЛУБЛЕННОГО ВЫСШЕГО ОБРАЗОВАНИЯ

Копыцкий А. В., Смирнов В. Ю., Левкович Т. В.

Гродненский государственный медицинский университет
Гродно, Беларусь

Актуальность. Построение моделей множественной регрессии на ограниченном числе выборочных наблюдений остается актуальной проблемой в прикладной статистике. Статистические модели необходимы для описания уравнений, отражающих связи между зависимыми переменными (откликами) и комбинациями независимых переменных (предикторов). Такие модели важны для интерпретации связей между реальными показателями, для прогнозирования с заданной точностью значений откликов при известных значениях предикторов. Регрессионные уравнения часто применяются в биомедицинской статистике при решении научных задач (в том числе и при написании диссертационных работ на уровнях углубленного и научно-ориентированного высшего образования), используются в инструкциях по применению, патентах на полезные модели, научных статьях и докладах. Для подбора оптимального сочетания предикторов в регрессионное уравнение используются разные приемы: методы пошагового включения/исключения, ручного подбора, отбор по одному предиктору с последующим их комбинированием, метод прямого перебора. Наиболее надежным является метод прямого перебора, однако он же требует наибольшего времени. Другие методы чувствительны к объемам выборок, к наличию пропущенных значений в исходных данных, и их применение для выборок ограниченного объема (часто встречающихся в биомедицинских исследованиях) затруднено. Прямой перебор моделей, которые могут быть построены на множестве возможных предикторов, как уже было указано выше, требует значительных вычислительных ресурсов, так как увеличение числа регрессоров ведет к быстрому росту числа моделей. Однако количество проверяемых моделей

может быть снижено за счет специфики биомедицинских исследований, исключения коррелирующих предикторов (что одновременно частично решает проблему мультиколлинеарности в моделях), включения в число возможных регрессоров заранее определенных переменных (представляющих особый интерес для исследователя). Кроме того, благодаря многопоточности вычислений в современных персональных компьютерах и их высокой производительности даже при большом числе моделей время их построения и фильтрации может быть вполне приемлемым (например, до 12 часов). Ранее нами было разработано программное решение для перебора и фильтрации регрессионных моделей, где реализованы указанные принципы [1]. В ходе эксплуатации программы-переборщика регрессионных моделей, тем не менее мы часто сталкивались с ситуациями, когда время перебора становилось довольно значительным (от нескольких суток). Поэтому стала актуальной модернизация нашего решения, позволяющая снизить указанное время. В качестве направления для модернизации был выбран метод итеративного отбора потенциальных предикторов.

Цель работы – модернизация программы-переборщика регрессионных моделей за счет итеративного метода отбора потенциальных предикторов. Для достижения указанной цели необходимо решение ряда задач:

- разработка и отладка модуля итеративного отбора потенциальных предикторов;
- разработка и отладка модуля для определения скорости перебора моделей и прогнозирования времени перебора;
- разработка и отладка модуля для принятия решения о включении итеративного метода;
- внедрение всех модулей в код и проверка работоспособности итогового программного решения.

Материалы и методы исследования. Решение описанных выше задач выполнено за счет анализа и рефакторинга кода программного решения, написанного ранее. Также разработаны и отлажены необходимые программные модули, после чего они были встроены в уже имеющийся программный код. Помимо того, выполнена оптимизация полученного решения, так как при наличии итераций некоторые процедуры (например, фильтрации сочетаний предикторов с отсеиванием пар коррелирующих регрессоров) выполнялись неоднократно, в чем уже не было необходимости. Кроме вышеописанного был разработан и имплементирован метод итерационного подбора возможных предикторов на основании частот их повторяемости в моделях, прошедших фильтрацию. Рассмотрим подробнее идею данного метода.

Пусть исходное множество потенциальных предикторов содержит N элементов. Без итерационной процедуры в нашей программе сначала строится множество из $N_1=N$ моделей с одним предиктором, после чего проводится их фильтрация. На следующем этапе анализируется множество из $N_2=N!/(N-2)!$ моделей с двумя регрессорами (на самом деле, N_2 может быть и меньше за счет предварительной фильтрации пар коррелирующих регрессоров либо за счет

принудительного включения необходимого регрессора во все модели). Таким образом, на этапе k необходимо построить и проанализировать N_k моделей, где

$$N_k = N! / (N - k)!,$$

а k последовательно принимает значения от 1 до M (M – максимальное число предикторов в модели). Так, например, при $N=60$ потенциальных предикторах и при максимальном числе предикторов $M=6$ получается 56 049 057 моделей. При средней скорости обработки в 400 моделей/с на их анализ уйдет 39 часов, и такое время ожидания может оказаться чрезмерным. Для снижения количества анализируемых моделей мы предлагаем итеративный метод сужения множества потенциальных предикторов. Суть его состоит в том, что в указанных выше формулах мы меняем число N на основании частоты встречаемости предикторов в моделях, прошедших фильтрацию на предыдущих этапах, то есть число моделей на k -м этапе с учетом рейтинга будет равно:

$$N_k' = N_{k-1}! / (N_{k-1} - k)!.$$

Для определения N_{k-1} строится частота встречаемости каждого предиктора в моделях, прошедших фильтрацию на этапах 1, 2, ..., $k-1$. Далее отбрасываются те предикторы, которые встречаются с частотой менее пороговой P_0 (например, менее 1%). Полученный набор из N_{k-1} потенциальных регрессоров будет использоваться в следующей k -й итерации перебора. Пусть в приведенном выше примере после 3-й итерации осталось не $N=60$ потенциальных предикторов, а, например, $N=40$. Тогда, если на всех последующих шагах число N остается равным 40, общее число моделей будет снижено до 4 623 828 и время перебора уменьшится до 3,2 часа.

С целью принятия решения о включении итеративного механизма для сужения множества потенциальных предикторов нами реализован модуль для оценки скорости обработки множества моделей, в котором определяется время, затраченное на анализ 10 000 моделей. На основании полученной скорости дальше оценивается время, необходимое на анализ всего множества моделей. Если это время превышает, например, 5 часов, то будет задействован итеративный метод.

Результаты. Описанный алгоритм имплементирован в программу-переборщик и апробирован на реальных биомедицинских данных, используемых в диссертационных исследованиях. Так, например, в исследовании факторов, провоцирующих старение крупных кровеносных сосудов (объем – 220 испытуемых, 50 полных наборов данных, для остальных 170 пациентов – данные с пропущенными значениями), при $N=61$ потенциальном предикторе, при максимальном числе регрессоров в уравнении $M=8$ и скорости анализа в 250 моделей/с прогнозируемое время перебора было равно 18 суткам. Однако при применении после 3-го шага перебора итеративного метода сужения набора потенциальных регрессоров (с отбрасыванием предикторов с частотой встречаемости в отфильтрованных моделях менее 1%) время полного анализа составило только 10 минут. Это связано с тем, что список потенциальных предикторов после 4-й итерации сократился до 15 элементов, а на последнем этапе – до 13.

Вывод. Таким образом, модифицированная программа-переборщик регрессионных моделей может использоваться для поиска оптимальных регрессионных уравнений, необходимых при анализе данных медико-биологических исследований.

Литература

1. Копыцкий, А. В. Информационно-вычислительная технология с применением языка «R» в рамках второй ступени высшего образования в медицинских вузах / А. В. Копыцкий. – Мн.: Вышэйшая школа. – 2021. – № 3. – С. 18-22.

ЦИФРОВЫЕ МЕТОДЫ В ОБУЧЕНИИ СТУДЕНТОВ ДИАГНОСТИЧЕСКОЙ СПЕЦИАЛЬНОСТИ ДЛЯ ОПРЕДЕЛЕНИЯ РАННИХ ПРИЗНАКОВ АУТОИММУННОЙ ОФТАЛЬМОПАТИИ

Кринец Ж. М.¹, Нечипоренко А. С.², Семянович Т. В.²

¹Гродненский государственный медицинский университет

²Гродненская университетская клиника

Гродно, Беларусь

Актуальность. Диагностика аутоиммунной офтальмопатии (АИО) основана на триаде классических признаков – дисфункция щитовидной железы, данные компьютерной томографии орбит (КТ) и наличие изменений со стороны глаз. КТ орбит позволяет визуализировать костные и мягкотканые структуры в трехмерной проекции, способствует выявлению ранних проявлений заболевания и позволяет определить риск развития экзофтальма.

Цель исследования. Определить основные компьютерно-томографические параметры костных структур орбиты и глазного яблока у пациентов с АИО.

Методы исследования. В отделении микрохирургии глаза УЗ «Гродненская университетская клиника» обследовано 48 пациентов (96 орбит) с активной АИО. Всем пациентам проводили КТ лицевого черепа в аксиальной плоскости на компьютерных томографах «Light Speed Pro 32» фирмы «GENERAL ELECTRIC» и «Aquilion Lightning TSX-036A». Постпроцессинговая обработка данных заключалась в построении мультипланарных реконструкций в корональной и сагиттальной плоскостях. Описательные статистики представлены в виде медианы 1-го и 3-го квартилей. Рассчитывали значение величины порога отсечения исследуемых параметров методом построения ROC-кривой и показателя площади под кривыми (AUC). Для получения унифицированной эмпирической формулы использованы методы высшей математики и прикладной статистики.

Результаты и их обсуждение. Возможность проведения числовых расчетов на изображениях орбит при КТ позволила определить начальные признаки АИО: экзофтальм, утолщенные (>4 мм) веретенообразной формы