

волевым потенциалом, плохим предвидением, прагматичные и с высокой внушаемостью со стороны) склонны отходить от своих основных жизненных задач, попадать в «опасные» компании и в итоге «формировать» событие смерти.

Резюмируя всё выше изложенное, отметим, что корректное статистическое исследование социальных объектов предполагает не только хорошее владение инструментом анализа данных, но и одновременно глубокое знание природы самих объектов.

Литература

1 Осипенко, К. А. Метод регрессионного моделирования продолжительности жизни по дате рождения / К. А. Осипенко, Н. Б. Осипенко // Творчество молодых 2012: сборник научных работ студентов и аспирантов УО «ГГУ им. Ф. Скорины»: в 2 ч. / Гомельский гос. ун-т им. Ф. Скорины; отв. ред. О. М. Демиденко. – Гомель, 2012. – Ч. 1. – С. 194–197.

УДК 519.25

К. А. Осипенко, А. Н. Осипенко

О ВОЗМОЖНОСТИ ИСПОЛЬЗОВАНИЯ ПАСПОРТНЫХ ДАННЫХ В СТАТИСТИЧЕСКОМ АНАЛИЗЕ ПРИЧИН СМЕРТНОСТИ НАСЕЛЕНИЯ

В статье описывается один из вариантов трехэтапного скрининга здоровья населения. Предлагается способ использования паспортных данных для предварительной экспресс-диагностики групп риска обследуемого человека. Даются алгоритмы перевода даты рождения, имени и фамилии в множество бинарных признаков для распознавания групп риска. Приводятся схема распознавания и формула дискриминантной функции. На примере пробного исследования показывается практическая целесообразность применения этих данных на первом этапе скрининга.

Обоснование целесообразности использования паспортных данных в статистическом анализе причин смертности населения

Задача выявления причин смертности и, в частности, особенностей основных групп риска по статистическим данным является одной из ведущих в сфере санитарно-гигиенических исследований. Особенность современного момента состоит в том, что появляются технические возможности для массового скрининга здоровья населения. Разрабатываются и внедряются различные концепции такого скрининга по поводу ранней диагностики и профилактики онкозаболеваний, туберкулеза, сердечнососудистых и других заболеваний. В связи с этим возникает проблема согласования и синхронизации всех этих обследований в рамках единой концепции общереспубликанского скрининга здоровья населения.

В настоящей работе предлагается для этих целей разработать методологию и программно-технологическое обеспечение предварительной экспресс-диагностики на основе паспортных сведений, данных о группе крови, антропометрии, анамнезе и иной стандартной информации, имеющейся в поликлинических базах данных. В перспективе к этой информации могут быть добавлены данные биометрии (отпечатки пальцев, фотографии сетчатки глаза), фрагменты почерка. В еще более далекой перспективе – данные недорогого экспресс-анализа ДНК.

В результате такой диагностики для каждого человека будут сформированы оценки вероятностей принадлежности к основным группам риска.

На втором этапе скрининга должно осуществляться подробное анкетирование граждан по поводу соответствующих целевых проблем в наиболее вероятных для них группах риска.

Наконец, только на третьем этапе предполагается перейти к более дорогостоящему детальному клиническому и амбулаторному обследованию (если в этом появляется

необходимость).

Наиболее спорным в предлагаемой трехэтапной схеме массового скрининга является первый этап. Основной довод – высокие ошибки диагностики. Как показывает проведенный нами пробный анализ статистических связей паспортных данных и основных причин смерти, такие связи имеются и они вполне пригодны для предварительного распределения жителей по группам риска.

Что касается ошибок этой диагностики, то, во-первых, окончательное решение об отнесении человека к группе риска (или к нескольким группам риска) на завершающей фазе первого этапа должен принимать участковый терапевт. Результаты компьютерной обработки по этому человеку носят статус дополнительной ориентирующей информации. Они позволят специалисту быстрее сфокусироваться на потенциально более слабых сторонах здоровья обследуемого. Особенно это касается молодежи и лиц среднего возраста, практически никогда самостоятельно не обращавшихся за медицинской помощью.

Во-вторых, ошибки первого этапа скрининга (отнесение не к своей группе риска) могут быть выявлены на втором этапе. В этой ситуации обследуемому придется провести дополнительное анкетирование по иной группе риска, подобранной для него врачом с учетом компьютерной обработки его первой анкеты. В любом случае, наличие первого этапа скрининга позволит уйти от обременительного анкетирования сразу по всем группам риска. В оптимальном варианте их должно быть не меньше десятка.

Описание исходных данных для статистического анализа

В качестве исходного материала для пробного исследования возможностей использования паспортных сведений на первом этапе скрининга послужила выборка из 88 знаменитых людей. По каждому человеку из интернета брались данные в виде строки: имя; фамилия; страна, в которой жил человек; тип менталитета; день, месяц и год рождения; день, месяц и год смерти; количество жен (мужей); количество детей (включая приемных); основной род занятий по жизни (профессия); общая причина смерти (заболевание сердечнососудистой системы – инфаркт или инсульт – 1, онкозаболевание – 2, другие болезни – 3, несчастные случаи – 4, преднамеренное убийство – 5, иное – 6); детализация подсистем организма в структуре общей причины смерти:

1 Причина смерти (сердечнососудистые заболевания):

- 1) сердце (инфаркт и др.);
- 2) мозг (инсульт и др.);
- 3) тромб и т. п.

2 Причина смерти (онкозаболевание):

- 1) нервная система (мозг и т. д.);
- 2) эндокринная (железы: щитовидная, поджелудочная и т. д.);
- 3) дыхательная;
- 4) пищеварительная;
- 5) мочева (почки и т. д.);
- 6) половая;
- 7) опорно-двигательная (кости);
- 8) кроветворная и иммунная;
- 9) органов чувств (глаза, кожа, уши, гортань, нос);

10) иное, включая: лимфа, селезенка, мягкие ткани, ткани сосудов.

3 Причина смерти (хронические болезни): детализация подсистем организма та же, что в (2) причине смерти (онкозаболевание).

4 Причина смерти (несчастные случаи):

- 1) авария;
- 2) смертельные травмы;
- 3) самоубийство;
- 4) отравление;
- 5) иное.

Для обработки данных использовался пакет «Statistica», в частности, его средства подготовки новых переменных путём того или иного преобразования исходных признаков, а также программы анализа вариантов, построения гистограмм, матриц корреляций признаков, классификации данных и дискриминантного анализа. Кроме того, в рамках этого пакета была написана программа обучения распознаванию групп риска.

Перевод даты рождения, имени и фамилии в качественные признаки

Перевод даты рождения в набор числовых качественных признаков осуществлен с помощью общеизвестного алгоритма Пифагора и описан в работе [1]. В итоге из даты рождения получаются следующие признаки: число (цифра) даты (получено путем поэтапного сложения всех ее цифр пока в сумме не получится одна цифра); количества встречаемости цифр 0, 1, ..., 9 в рабочих числах алгоритма Пифагора [1]; образованные из предыдущих десяти номинальных признаков бинарные признаки, показывающие наличие или отсутствие, например, ноль двоек, одной двойки и т. д. в квадрате Пифагора.

Для перевода имени и фамилии в признаки для распознавания групп риска нумерология предлагает множество различных соответствий букв и цифр. В настоящей работе использована числовая азбука [2], с помощью которой можно вместо буквы ставить числа: А-1, Б-2, В-3, Г-4, Д-5, Е-6, Ж-7, З-8, И-9, К-10, Л-20, М-30, Н-40, О-50, П-60, Р-70, С-80, Т-90, У-100, Ф-200, Х-300, Ц-400, Ч-500, Ш-600, Щ-700, Ю-800, Я-900, Э-1000. В работе [2] приведена таблица содержательных значений этих числовых кодов. Например, для известного баснописца Ивана Крылова кодирование имеет следующий вид: И-9, В-3, А-1, Н-40. В сумме 53; К-10, Р-70, Ы-0, Л-20, О-50, В-3.

В сумме 153. Складываем обе полученные суммы ($53 + 153 = 206$). Полученное число ищем в приведенной в [2] таблице. Но так как точно числа 206 в ней нет, то его следует разбить на $200 + 6$. Для числа 200 находим: хладнокровие, слабохарактерность, для 6 – труд, свободолюбие, успех. Как видно из биографии знаменитого баснописца, все эти качества были налицо. Теперь поступим с числом 206 иначе. Сложим его цифры до двузначного или однозначного вида. И снова обратимся к таблице содержательной интерпретации чисел: 8 – величие, кротость, справедливость.

При наличии обучающей выборки в несколько сотен тысяч человек можно было бы воспользоваться полностью описанной здесь кодировкой. В нашем же пробном примере применялась только однозначная кодировка. При этом были образованы группы бинарных признаков для имени, для фамилии и для имени вместе с фамилией. Соответствующие числа (цифры) преобразовывались в последовательность из девяти бинарных признаков. Например, у семерки на седьмом месте в такой последовательности стоит единица, а на других местах – нули.

Алгоритм распознавания групп риска

В связи с небольшим объемом пробной выборки вместо четырех общих групп риска были выделены две. В первую группу вошли умершие от сердечнососудистых заболеваний и

от несчастных случаев. Для нее характерна быстрая потеря жизнеспособности организма. Во вторую группу вошли те, у кого такая потеря шла постепенно – это умершие от онкозаболеваний и других хронических болезней. Также из-за недостаточного объема исходной выборки не удалось осуществить полноценное выделение ее однородных подвыборок, в частности, использовать для этого признаки пола, менталитета и продолжительности жизни (ранние смерти и смерти в зрелом возрасте). Сам алгоритм распознавания групп риска разбит на два этапа.

Этап 1. Обучение распознаванию групп риска.

1. Выбор информативных признаков.

1.1. Построение матрицы взаимных корреляций по всем признакам, включая целевой признак принадлежности к первой или второй обобщенным группам риска.

1.2. Выбор признаков, имеющих более или менее значимую связь с целевым свойством (в качестве критерия выбора информативного признака в нашем случае бралось условие: $|r| \geq 0,1$). Положительная корреляция говорит о том, что признак в большей степени «голосует» за вторую группу риска и, наоборот, отрицательная корреляция говорит о предпочтении первой группы риска.

1.3. Формирование дискриминантной функции.

Пусть $P_1 = \{p_1^1, p_2^1, \dots, p_{n_1}^1\}$ – множество признаков, голосующих за первую группу риска, а $P_2 = \{p_1^2, p_2^2, \dots, p_{n_2}^2\}$ – множество признаков, голосующих за вторую группу риска. Вероятность B_1 отнесения объекта x к классу 1 вычисляется по формуле:

$$B_1 = (\sum_{i=1}^{n_1} x_i^1 / n_1) / ((\sum_{i=1}^{n_1} x_i^1) / n_1 + (\sum_{j=1}^{n_2} x_j^2) / n_2),$$

где $x_i^1, (i = 1, \dots, n_1)$ – значения признаков из множества P_1 для объекта x ;

$x_j^2, (j = 1, \dots, n_2)$ – значения признаков из множества P_2 для объекта x .

Соответственно, $B_2 = 1 - B_1$ – вероятность отнесения объекта x к классу 2.

Этап 2. Экзамен алгоритма распознавания групп риска.

Обычно экзамен проводится для объектов, не участвовавших при построении дискриминантной функции. Принятие решения об отнесении объекта к группе риска имеет вид:

$$R = \begin{cases} 1, & \text{если } B_1 - B_2 \geq \alpha; \\ 2, & \text{если } B_2 - B_1 \geq \alpha; \\ 0, & \text{если } |B_1 - B_2| < \alpha, \end{cases}$$

где $0 \leq \alpha \leq 0.5$ (в нашем примере $\alpha = 0.2$).

Если $R = 0$, то программа не может различить группу риска для объекта x .

Результаты распознавания групп риска

Оценка качества дискриминантной функции проводилась по исходной выборке. При этом ошибка первого рода (отнесение объекта первого класса ко второму) составила 10/55, ошибка второго рода – 1/33. Доля отказов – 39/88.

Высокую ошибку распознавания первой группы (в основном это смерти от сердечнососудистой болезни) можно объяснить неполнотой диагноза смертельного исхода, особенно в зрелом возрасте, когда инфаркт или инсульт становится следствием какой-либо хронической болезни. Часто это не афишируется даже для знаменитостей и никак не отражается в диагнозах смерти.

Заключение

Несмотря на небольшой объем выборки в проведенном пробном исследовании, можно утверждать, что паспортные данные вполне пригодны для включения их в список признаков при экспресс-диагностике группы риска на первом этапе скрининга здоровья населения.

Литература

- 1 Осипенко, К. А. Метод регрессионного моделирования продолжительности жизни по дате рождения / К. А. Осипенко, Н. Б. Осипенко // Творчество молодых 2012: сборник научных работ студентов и аспирантов УО «ГГУ им. Ф. Скорины»: в 2 ч. / Гомельский гос. ун-т им. Ф. Скорины; отв. ред. О. М. Демиденко. – Гомель, 2012. – Ч. 1. – С.194–197.
- 2 Хигир, Б. Ю. Число имени / Б. Ю. Хигир. –СПб.: Астрель, 2008. – 42 с.

УДК 546.26

А. И. Остапенко

ИССЛЕДОВАНИЕ ЭНДОЭДРАЛЬНЫХ КОМПЛЕКСОВ ФУЛЛЕРЕНА C₆₀ С ПОМОЩЬЮ ПРОГРАММНОГО ПАКЕТА HYPERCHEM

В статье изложены результаты расчетов энергии образования и оптимальной геометрии эндоэдральных комплексов фуллерена C₆₀ и металлофуллеренов, полученные с применением программного пакета HyperChem.

Свойства нано-размерных систем широко исследуются в настоящее время как экспериментальными и теоретическими методами, так и методами вычислительного эксперимента. Наиболее интересными объектами в данных исследованиях являются фуллерены [1].

Обнаружение в 1985 году Крото углеродного кластера C₆₀ при изучении масс-спектров паров графита, образующихся под действием лазерного излучения, послужило началом исследования фуллеренов – аллотропных кристаллических модификаций углерода, построенных из полых сфероидальных молекул C_n. Следует отметить, что существование подобных структур было предсказано еще ранее теоретически. Так, Осаво самостоятельно в 1970 году, а затем в соавторстве с Йошида в 1971 обсуждал возможность существования трехмерных икосаэдрических структур, состоящих из атомов углерода. Они дали графическое изображение структуры, соответствующей фуллерену C₆₀, в образе футбольного мяча [1].

Расчет свойств таких структур удобно реализовывать в пакетах квантово-химического моделирования. Одним из самых известных является HyperChem. Он может выполнять расчеты энергии систем и их равновесной геометрии методом молекулярной механики, девятью полуэмпирическими квантово-химическими методами (Расширенный метод Хюккеля, CNDO, INDO, MINDO3, MNDO, AM1, PM3, ZINDO/1 и ZINDO/S), или теоретическим (ab initio) методом квантовой химии в различных базисах [2].

В методе молекулярной механики атомы рассматриваются как ньютоновские частицы, которые взаимодействуют друг с другом посредством потенциальных полей, задаваемых эмпирически. Потенциальная энергия взаимодействия зависит от длины связей, углов связи, торсионных углов и нековалентных взаимодействий (в том числе сил Ван-дер-Ваальса, электростатических взаимодействий и водородных связей). В этих расчетах силы, действующие на атомы, представляются в виде функций координат атомов.

Полуэмпирические методы решают уравнение Шредингера для атомов и молекул с использованием определенных приближений и упрощений. Все методы этой группы характеризуются следующими особенностями: расчет ведется только для валентных электронов, используются стандартные неоптимизированные базисные функции электронных орбиталей и некоторые экспериментальные параметры. С помощью экспериментальных параметров устраняют необходимость расчетов ряда величин и корректируют ошибочные результаты приближений.

Метод AM1 является одним из наиболее точных методов. Он используется для молекул, содержащих элементы из главных подгрупп 1 и 2 групп периодической системы. С его помощью