

*Суворова Алёна Владимировна,
Тулупьев Александр Львович,
Пашенко Антон Евгеньевич,
Тулупьева Татьяна Валентиновна,
Красносельских Татьяна Валерьевна*

УДК 311.2 + 616-036.22

АНАЛИЗ ГРАНУЛЯРНЫХ ДАННЫХ И ЗНАНИЙ В ЗАДАЧАХ ИССЛЕДОВАНИЯ СОЦИАЛЬНО ЗНАЧИМЫХ ВИДОВ ПОВЕДЕНИЯ

Аннотация

В статье рассматриваются различные прикладные задачи, связанные с необходимостью оценки параметров поведения определенных групп на основе гранулярных данных (и более широко – гранулярных знаний). Кроме того, описываются проблемы, возникающие при попытке представить и обработать такие данные и знания в интеллектуальных системах, обозначены возможные пути решения указанных проблем. Также в статье перечисляются исследовательские задачи, стоящие в рассматриваемой области.

Ключевые слова: дефицит информации, гранулярность данных, знания с неопределенностью, модели поведения, сверхкороткие неточные ряды, оценки риска, оценки интенсивности.

Одной из задач маркетинговых исследований является поиск ответа на вопрос, в каком объеме члены определенных групп потребляют товары (пользуются услугами) за заданный период времени в конкретном месте – это позволяет выделить группы потребителей, существенно различающиеся интенсивностью потребления продуктов, товаров или услуг. При наличии таких результатов маркетинговые усилия можно сосредоточить на тех группах, которые многочисленны, но товар потребляют неинтенсивно. Выявление их особенностей может привести к разработке стратегии,

ведущей к существенному увеличению объема продаж.

Следует отметить, что необходимые данные об интенсивности потребления невозможно получить из анализа продаж, то есть недостаточно изучить «чеки» – данные о состоявшихся продажах. Это дает возможность принять во внимание лишь те группы, которые и так уже покупают данный товар. Не исключено, что в таком случае из анализа выпадут многочисленные потенциальные потребители, которые ни разу еще не употребляли интересующие маркетологов товары или услуги и которые как раз и составляют «неосвоенную» нишу для продаж.

Под интенсивностью поведения понимается число его эпизодов в определенный промежуток времени. Существует несколько

© А.В. Суворова, А.Л. Тулупьев,
А.Е. Пашенко, Т.В. Тулупьева,
Т.В. Красносельских, 2010

ко подходов, которые позволяют регистрировать эпизоды поведения различного вида. В частности, «дневниковый метод» подразумевает запись всех действий респондента (испытуемого, клиента, пациента) в течение дня, после чего полученные данные, как правило, накапливавшиеся несколько месяцев, поступают на обработку эксперту, который подсчитывает число эпизодов поведения определенного вида за данный период. Однако такой вид исследований достаточно дорог, его сложно организовать и долго выполнять. Вместо него число эпизодов можно оценить, если известна оценка интенсивности поведения, рассмотренного как случайный процесс определенного класса. При этом встает задача оценки интенсивности рассматриваемого поведения респондента по его «одномоментному» самоотчету, то есть по ответам в анкете на блок вопросов или по результатам проведения интервью. Заметим, что подобные опросы опираются на информацию, хранящуюся в памяти респондента, и, естественно, чем глубже ретроспектива, тем труднее респондентам отвечать на вопросы и тем больше они делают ошибок припоминания.

Сейчас для получения такого самоотчета используются два метода, каждый из которых имеет недостатки [1]. Первый метод – прямые вопросы: «Сколько раз Вы *делали так* в течение последнего месяца (трех, шести, года)?». На такие вопросы респонденты обычно дают практически не соответствующие с реальностью ответы. Действительно, можно задать себе вопрос: «Сколько раз за последние три месяца я пил чай с сахаром?». Попытка ответа даже самому себе даст четкую картину незначительной достоверности такого ответа.

Второй метод – Лайкерт-шкалы – опросники, в которых используются качественные, а не количественные варианты: «Никогда», «Редко», «Иногда», «Часто», «Всегда» и подобные им возможности для ответа. Вопрос ставится легко, ответ тоже получить несложно, однако эти ответы не несут никаких полезных сведений относи-

тельно числа эпизодов¹: то, что «Часто» для одного человека, может быть «Редко» для другого, а то, что «Часто» в одном виде поведения, может быть «Редко» для другого вида поведения. Кроме того, «расстояние» между «Всегда» и «Очень часто» совершенно не обязательно совпадает с расстоянием между «Редко» и «Никогда». На практике шкалы арифметизируют, но за этой арифметизацией не стоит никакой достоверной гипотезы; получающиеся расчеты ситуацию с интенсивностью поведения не характеризуют вообще никак. Таким образом, возникает потребность в более адекватных источниках сведений о социально значимом поведении и методиках их обработки, которые сделают возможной более обоснованную оценку числа эпизодов.

Одной из возможных альтернатив Лайкерт-шкал представляется опрос респондента об одном или нескольких последних эпизодах его поведения. Такой опрос позволяет судить об интервалах между эпизодами рискованного поведения, а также об интервале между временем опроса и последним эпизодом.

Уже сами упомянутые интервалы могут оказаться более удобными косвенными оценками риска, чем порядковые шкалы вида «Никогда, Редко, Иногда, Часто, Всегда». Чем меньше интервалы, тем более высокую степень риска мы можем ожидать. Интервалы можно сравнивать между собой. Существуют еще ряд преимуществ данного подхода [1]:

- ответы не надо арифметизировать, они и так количественные, более того – континуальные;

- интервалы можно сравнивать между собой;

- можно рассчитывать коэффициенты корреляции – данный инструмент наиболее (хотя и не всегда вполне обоснованно) востребован в гуманитарных науках для поиска зависимостей как между ответами на отдельные вопросы, так и на группы вопросов.

Ответы о последних эпизодах обладают определенной особенностью: их формули-

¹ Попробуйте на основе ответов «Всегда, Часто, Иногда...» о потреблении того или иного товара *количественно* оценить объем продукции, который надо подготовить к продажам.

ровки *гранулярны*, то есть бытовой язык диктует использование привычных оценок длительности, не отличающихся особой точностью. Например, высказывание «неделю назад» не означает то же самое, что «семь дней назад» или «168 часов назад». Гранулярность данных приходится учитывать, речь о ней пойдет ниже.

Отметим, что задача оценки интенсивности поведения стоит и в других отраслях прикладных исследований. Одной из них является медицина. Например, при одном из самых распространенных заболеваний эндокринной системы – сахарном диабете, – в основе которого лежит относительный или абсолютный дефицит инсулина, важнейшее значение имеет строгое соблюдение диеты. В некоторых случаях при диабете 2-го типа для компенсации нарушения углеводного обмена и прекращения прогрессирования заболевания достаточно лишь ограничить употребление легкоусвояемых углеводов и жиров. При 1-м типе сахарного диабета соблюдение диеты жизненно важно для больного, ее нарушение может привести к гипо- или гипергликемической коме, а иногда – к смерти. Лечащему врачу необходимо оценивать частоту и обстоятельства отклонения пациента от диеты, чтобы иметь возможность корректировать дозировку назначаемых пероральных сахароснижающих препаратов или инсулинов, а также понимать, какие вмешательства необходимо предпринять для повышения приверженности больного рекомендациям по соблюдению диеты. Таким образом, выбор тактики ведения больного сахарным диабетом во многом основывается на степени интенсивности отклонения от диеты.

Приведем другой пример. Многочисленные исследования свидетельствуют о значительных нарушениях социально-психологической адаптации ветеранов боевых действий, которые в равной степени составляют как медицинскую, так и социальную проблему. Характерными особенностями социального поведения участников боевых действий в отдаленном периоде боевого психического стресса являются склонность к дисфориям, вспыльчивость, приступы гнев

ва и неконтролируемая агрессивность по отношению как к окружающим, так и к себе. Агрессивность играет существенную роль в возникновении социальной дизадаптации, в развитии психосоматических и пограничных нервно-психических расстройств, а также в саморазрушающем поведении участников войн. Изучение частоты эпизодов и степени тяжести проявлений агрессивного поведения, то есть оценка интенсивности агрессии, позволяет врачам и социальным психологам оценивать уровень социальной адаптации, эффективность лечения, психотерапии и коррекции [2].

Еще одним примером является изучение поведения, рискованного в отношении заражения инфекциями, передаваемыми половым путем (ИППП), в эпидемиологии [1, 3–14]. В настоящее время наиболее острой эпидемиологической проблемой является оценка риска передачи и приобретения такой опасной и неизлечимой инфекции, как инфекция вирусом иммунодефицита человека (ВИЧ) в зависимости от особенностей инъекционного и сексуального поведения индивида. Наиболее точно такой риск характеризуется инцидент-показателем – числом заразившихся за определенный период среди лиц, подвергавшихся риску заражения, отнесенным к человеко×месяцам наблюдения. Для прямого измерения инцидент-показателя требуется организовать когортное исследование, подразумевающее вовлечение, как правило, не менее 500 представителей группы риска и их медицинское и социальное сопровождение в течение значительного периода времени. Однократное проведение подобного когортного исследования обходится в полтора-два миллиона долларов. Такой уровень расходов делает затруднительным мониторинг инцидент-показателя даже в странах с сильной экономикой. Требуется предложить математические модели, позволяющие выполнить более дешевые косвенные измерения инцидент-показателя на основе ответов респондентов, составляющих выборку из группы риска. Один из таких способов опирается на модель Белла-Тревино [16]. Инцидент-показатель можно оце

нить, зная индивидуальный риск заражения за заданный период времени каждого отдельного респондента. Модель Белла-Тревино увязывает оценку риска с числом эпизодов рискованного поведения. Число же эпизодов можно оценить, если, в свою очередь, известна оценка интенсивности рискованного поведения, рассмотренного как случайный процесс определенного класса.

Цель данной статьи – на примере приведенных выше прикладных задач измерения параметров поведения ознакомить читателя с проблемами, возникающими при попытке представить и обработать гранулярные данные (или более широко – гранулярные знания) в интеллектуальных системах, а также обозначить возможные пути решения таких проблем. Кроме того, в статье перечисляются исследовательские задачи, стоящие в рассматриваемой области. Многие такие задачи связаны с тем, что оценки параметров приходится получать в условиях информационного дефицита, проявлениями которого и являются гранулярность данных, их неточность, неполнота и нечисловой характер.

Для удобства дальнейшего изложения остановимся на наиболее социально значимой из приведенных выше проблем – оценке риска инфицирования ВИЧ в зависимости от интенсивности рискованного поведения.

Уже отмечалось, что в случае опроса респондентов об особенностях их поведения данные поступают на естественном языке, то есть являются в значительной степени нечеткими и неполными. Такие высказывания необходимо систематизировать, классифицировать и формализовать для их последующей обработки. Ограниченное число и неточность, неопределенность, нечеткость естественно-языковых формулировок ответов не позволяют напрямую использовать известные методы из теории массового обслуживания для оценки интенсивности поведения.

Отметим, что респонденты используют в своих высказываниях преимущественно следующие единицы измерения: часы, дни, неде-

ли, месяцы, полугодия, года. Причем использованная единица измерения несет в себе информацию о точности измерения. Поясним это на примере двух высказываний: «семь дней назад» и «неделю назад». Когда респондент использует формулировку «семь дней назад», это свидетельствует о высокой «надежности» припоминания и его уверенности в том, что событие произошло ровно семь дней назад. Когда респондент использует формулировку «неделя назад», он априорно снижает точность высказывания. Неделя назад – это и шесть дней назад и восемь. Таким образом, можно говорить о гранулярности получаемых ответов (рис. 1). На рис. 1 схематично представлены сведения о нескольких последних эпизодах поведения. Пусть (0) – это момент интервью, (1), (2), (3) – моменты на оси времени, когда произошел последний, предпоследний и пред-предпоследний эпизод поведения. t_{01} , t_{12} , t_{23} – длины временных интервалов соответственно между моментом интервью и последним эпизодом, последним и вторым эпизодом, вторым и третьим эпизодом поведения в прошлом, t – весь временной промежуток, за который произошли рассматриваемые эпизоды.

Респондент не готов высказаться более точно, чем это определено «размерами» гранул. Если в опросе настаивать на указании более точных ответов, например, просить отметить конкретную дату, то можно получить формально более точные, а фактически не соотносящиеся с действительностью результаты.

В случае ВИЧ-рискованного поведения были выявлены три класса ответов [4–6]:

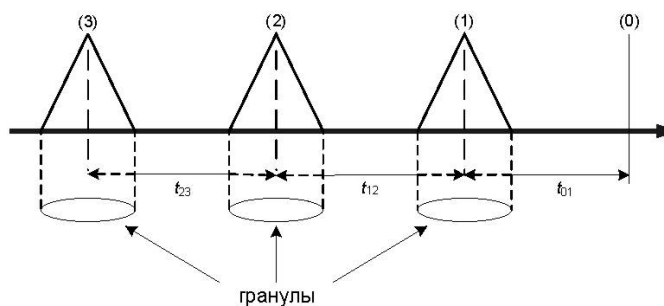


Рис. 1. Гранулярность ответов

– *вложенные интервалы*, когда указывается временной интервал между моментом интервью и каждым эпизодом, например: «вчера, позавчера, четыре дня назад» (рис. 2);

– *последовательные интервалы*, когда респондент указывает эпизоды своего поведения, начиная с предпоследнего, отсчитывая их от момента предыдущего эпизода, например: «вчера, за неделю до этого, за неделю до этого» (рис. 1, в случае (1) – «вчера», (2) – «за неделю до этого», (3) – «за неделю до этого»);

– *смешанные интервалы*, являющиеся комбинацией предыдущих двух классов, например: «вчера, позавчера, еще за день до этого».

Для обработки последнего класса используется обобщенная схема представления ответов. В этом случае эксперт имеет дело с объединенными в последовательности эпизодами поведения, для каждой последовательности дается оценка, а итоговая оценка складывается из суммы оценок таких последовательностей.

Эпидемиологам требуется определить или оценить величину параметра λ , характеризующего интенсивность участия респондента в поведении определенного вида, которое описывается пуассоновским случайным процессом (не следует упускать из виду, что возможно использование и других моделей). Получив оценку параметра λ , можно посчитать вероятность того, что в интервале $[t_0, t_0 + t]$ произойдут k событий:

$$P[N([t_0, t_0 + t]) = k] = \frac{e^{-\lambda t} (\lambda t)^k}{k!}.$$

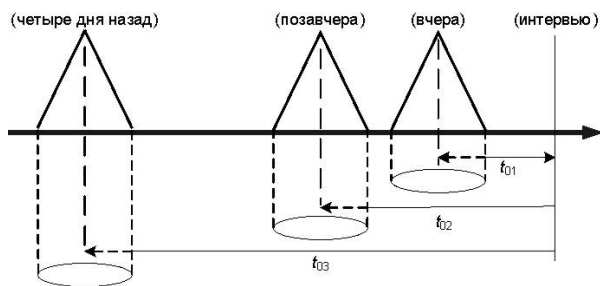


Рис. 2. Вложенные интервалы

Один из вариантов получения оценки интенсивности был рассмотрен в [7]. Пусть дано v – число последовательных эпизодов от момента интервью, которые вспомнил респондент, а $\tau = t_{0v}$ – тот период времени, за который эти эпизоды произошли. Применим метод максимального правдоподобия к основному уравнению пуассоновского процесса при вышеуказанных данных, чтобы найти соответствующую оценку интенсивности λ :

$$g(\lambda) = \frac{(\tau\lambda)^v}{v!} e^{-\tau\lambda},$$

$$h(\lambda) = \ln g(\lambda) = v \ln \lambda + v \ln \tau - \ln v! - \tau\lambda,$$

$$\frac{dh(\lambda)}{d\lambda} = \frac{v}{\lambda} - \tau,$$

$$\frac{dh(\lambda)}{d\lambda} = 0 \Rightarrow \lambda = \frac{v}{\tau}.$$

Как правило, исходя из ответов респондентов, удается дать численную оценку числу произошедших эпизодов между моментом интервью и наиболее отдаленным эпизодом включительно. Если, в частности, респондент ответил на все вопросы о последних трех эпизодах, то $v = 3$.

В силу существенной неопределенности высказываний на естественном языке, получить точную численную оценку τ (в рассматриваемом случае τ – это величина временного интервала между моментом интервью и самым отдаленным от него эпизодом включительно) затруднительно или даже невозможно. Однако ее можно рассмотреть как случайную величину, построенную над другими случайными величинами. Особенности процесса построения такой случайной величины подробно рассмотрены в [7].

Пусть известны данные о N последних эпизодах поведения, $t_1, t_2, t_3, \dots, t_N$, а τ – общий временной промежуток, за который произошли эпизоды (связь между t_1 и τ подробно рассмотрена в [1]). Интенсивность поведения λ оценивается по формуле: $\lambda = \frac{N}{\tau}$.

Для каждого эпизода со значением t_i , $1 \leq i \leq N$ (N – число рассматриваемых эпизодов поведения) через характеристику разброса δ определяется ин-

тервал (возможных значений) в днях: $[t_i - \delta x, t_i + \delta x]$, где x – коэффициент перевода рассматриваемой единицы измерения в дни [7].

То есть оценка в днях включает в себя минимальное значение $t_i - \delta x$, указанное в интервью значение t_i и максимальное значение $t_i + \delta x$.

Заметим, что любая точка из интервала $[t_i - \delta x, t_i + \delta x]$ возможна в качестве значения оценки t_i , что, однако, не означает, что точки из этого интервала равновероятны в качестве такого значения.

Сведения о такого рода отношениях между допустимыми значениями можно задать с помощью их распределения вероятностей. В зависимости от предположений о характере ответов респондента для задания случайной величины \hat{t}_i оценки t_i , используется равномерное, биномиальное или какое-либо другое вероятностное распределение.

Введенная случайная величина \hat{t}_i за счет рандомизации неопределенности ответа позволяет рассмотреть интенсивность как случайную величину и вычислить характеристики последней.

Для каждого эпизода соответствующий интервал разбивается на n частей. Рассматриваются все возможные сочетания точек из интервалов, соответствующих указанным эпизодам. Например, если исследуемые интервалы не пересекаются, то возможными являются все сочетания точек (рис. 3а). В случае же пересечения интервалов из рассмотрения исключаются не согласованные сочетания. На рис. 3б при выборе из предпоследнего (левого) интервала пятой точки (выделена окружностью) из последнего (правого) невозможно выбрать первую и вторую точки – иначе получится, что последний эпизод произошел раньше предпоследнего.

Расчет среднего значения для случая трех последних эпизодов производится по следующей формуле:

$$\lambda_{\text{среднее}} = \sum_{i,j,k} (\lambda_{ijk} p_i p_j p_k),$$

где p_i – вес i -й точки из первого интервала, p_j – вес j -й точки из первого интервала,

p_k – вес k -й точки из первого интервала, λ_{ijk} – оценка интенсивности для соответствующего сочетания точек, то есть $\lambda_{ijk} = N/\tau_{ijk}$, где τ_{ijk} – соответствующая точкам i, j, k оценка величины рассматриваемого интервала.

Частные случаи этой формулы для конкретных распределений подробно рассмотрены в [4]. Например, для равномерного

$$\lambda_{\text{среднее}} = \frac{1}{n^3} \sum_{i,j,k} \lambda_{ijk}.$$

Особый интерес представляют ответы респондентов, содержащие сведения о максимальном и минимальном интервале между эпизодами рассматриваемого поведения за заданный период времени. Оценка интенсивности для рискованного поведения на основе этих данных получена в [6, 7]. Отметим, что моделью поведения выступает пуассоновский процесс с уравнением

$$\Pr(\Delta t, k, \lambda) = \frac{(\lambda \Delta t)^k}{k!} e^{-\lambda \Delta t},$$

где Δt – промежуток времени наблюдения за поведением респондента; k – число эпизодов рассматриваемого поведения, случившихся в этот промежуток; λ – интенсивность поведения; $\Pr(\Delta t, k, \lambda)$ – вероятность того, что за промежуток времени наблюдения Δt при поведении с интенсивностью λ случится ровно k эпизодов указанного поведения.

Заметим, что для такого процесса известна также плотность распределения T –

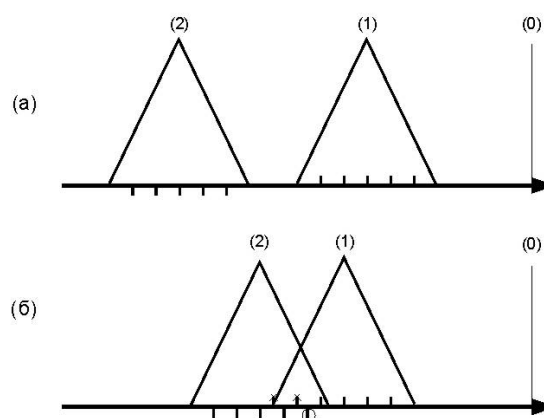


Рис. 3. Возможные сочетания точек

длины временного интервала между двумя соседними эпизодами: $p(T) = \lambda e^{-\lambda T}$, $T \geq 0$.

Пусть заданы T_{\max} – максимальная длина временного интервала между двумя соседними эпизодами и T_{\min} – минимальная длина временного интервала между двумя соседними эпизодами. Для того чтобы получить оценку интенсивности λ по этим данным, рассмотрим эвристический (опирающийся на некоторую рациональную идею, но еще не нашедший строгого математического обоснования) подход, нацеленный на максимизацию вероятности того, что длины интервалов между эпизодами попадают в промежуток $[T_{\min}; T_{\max}]$.

Такая оценка $\hat{\lambda}$, основанная на известных экстремальных значениях, вычисляется по формуле:

$$\hat{\lambda} = \frac{\ln \frac{T_{\max}}{T_{\min}}}{T_{\max} - T_{\min}},$$

однако ее поведение и свойства с математической точки зрения еще только предстоит исследовать.

Развитие описанных методов, рассмотрение возможных вариантов исходных данных, систематический учет их особенностей и применение приемов, использующихся в мягких вычислениях, гранулярных вычислениях, теории нечетких рядов, теории вероятностей и математической статистике, получение агрегированной оценки всех доступных данных в различных комбинациях, а также интервальной оценки для частоты поведения или числа эпизодов относятся к возможным и отличающимся особой актуальностью направлениям дальнейших исследований. Например, подход к оцениванию интенсивности поведения на основе данных о минимальном и максимальном интервалах между его эпизодами, связанный с использованием порядковых статистик, рассматривается в [8].

Кроме того, на данный момент остался не полностью рассмотренным вопрос о согласованности различных исходных данных [12], например, о согласованности данных о последних эпизодах рискованного поведения между собой, о согласованности дан-

ных о «рекордных» (минимальном и максимальном) интервалах между эпизодами рискованного поведения и, наконец, о согласованности данных о последних эпизодах и данных о «рекордных» интервалах. Зачастую респонденты, особенно люди со сложным социальным положением, хотят, чтобы у интервьюера сложилось о них положительное впечатление, и сознательно дают социально ожидаемые ответы; иногда они могут запутаться или ошибиться при ответе. Данная проблема, на самом деле, является одной из наиболее острых при проведении исследования.

Рассмотрим одну из задач подробнее. Респондент дает ответы как о последних эпизодах своего поведения, так и максимальном и минимальном интервалах между эпизодами того же поведения. По каждому типу таких данных можно вычислить свою оценку интенсивности, вопрос в том, как они будут соотноситься друг с другом, ведь они должны описывать одно и то же поведение. Таким образом, нужно предложить меру, позволяющую, например, по оценке, вычисленной по данным о последних эпизодах поведения, определить правдоподобность указанных респондентов минимального и максимального интервала между эпизодами его поведения.

Еще одно направление дальнейшего исследования – построение относительных оценок интенсивностей и рисков (так называемых odds ratio) [6, 9]. Такие оценки необходимы для сравнения разных групп или же для сравнения характеристик одной и той же группы до и после проведения поведенческой интервенции – комплекса мероприятий, нацеленного на модификацию поведения, например на снижение интенсивности участия в поведении, которое связано с угрозой приобретения или передачи гепатита С (под интенсивностью понимается число эпизодов поведения рассматриваемого вида в определенный промежуток времени). Требуется оценить *априорные* (до вмешательства) и *апостериорные* (после вмешательства) оценки интенсивности и кумулятивного риска, а также отношения этих оценок, характеризующие

влияние поведенческой интервенции на индивидов. То есть нужно по двум наборам данных — до вмешательства и после него — определить, изменилась ли интенсивность поведения. Понятно, что если до вмешательства происходило три эпизода в месяц, а после — три в неделю, то интенсивность поведения увеличилась, но такое явное изменение встречается очень редко. Необходимо разработать методы, позволяющие оценить влияние небольшого изменения исходных данных на изменение интенсивности.

Не менее важна и проблема автоматизации опроса, для чего требуется разработать удобный пользовательский интерфейс, учитывающий особенности респондентов.

Изучение социально значимого поведения — комплексная проблема, которая дол-

жна решаться на междисциплинарном уровне с привлечением многих наук о человеке, в первую очередь, медицинской и социальной психологии, что позволит оценить ее взаимосвязи с другими аспектами и сферами социальной жизни индивидов, их личностными и клиническими характеристиками. Трудно переоценить также значимость для клинической эпидемиологии тех огромных преимуществ и возможностей, которые предоставляет использование инструментов математики и биостатистики. Предлагаемый в настоящей статье подход к представлению и анализу гранулярных данных, основанный на данных о серии последних эпизодов социально значимого поведения, представляется авторам перспективным для использования в поведенческих исследованиях самой разной направленности.

Литература

1. Пащенко А.Е., Тулупьев А.Л., Николенко С.И. Статистическая оценка вероятности заражения ВИЧ-инфекцией на основе данных о последних эпизодах рискованного поведения // Труды СПИИРАН. Вып. 3, т. 1. СПб.: Наука, 2006.
2. Колов С.А., Остапенко А.В., Давыдова Е.В. Оценка агрессивного поведения как фактора социальной дизадаптации у ветеранов боевых действий. Бюллетень Волгоградского научного центра РАМН. 2009. № 1. С. 9–12.
3. Пащенко А.Е. Идентификация интенсивности пуассоновского процесса, моделирующего поведение респондента, в условиях дефицита информации. Информационно-измерительные и управляющие системы. 2009. № 4. т. 7. С. 45–48.
4. Пащенко А.Е., Суворова А.В. Программный комплекс для экспертного оценивания интенсивности поведения респондента в условиях дефицита информации // Интегрированные модели, мягкие вычисления, вероятностные системы и комплексы программ в искусственном интеллекте. Научно-практическая конференция студентов, аспирантов, молодых ученых и специалистов (Коломна, 26–27 мая 2009 г.). Научные доклады. В 2-х т. Т. 2. М.: Физматлит, 2009. С. 220–241.
5. Пащенко А.Е., Тулупьева Т.В. Применение процедуры рандомизации для оценки интенсивности поведения респондента в условиях информационного дефицита // Интегрированные модели и мягкие вычисления в искусственном интеллекте. V-я Международная научно-практическая конференция. Сборник научных трудов. В 2-х т. Т. 1. С. 743–751.
6. Тулупьев А.Л., Суворова А.В., Тулупьева Т.В., Пащенко А.Е. Косвенные оценки и сравнение параметров угрожающего поведения в разных группах по неполным и неточным данным // Международная конференция по мягким вычислениям и измерениям. Сборник докладов. 2009. Т. 2. СПб.: Изд-во СПбГЭТУ «ЛЭТИ», 2009. С. 110–114.
7. Пащенко А.Е., Тулупьев А.Л., Тулупьева Т.В. Оценка интенсивности рискованного поведения на основе нечетких ответов респондентов // Нечеткие системы и мягкие вычисления (НСМВ-2008): Сборник научных трудов Второй Всероссийской научной конференции с международным участием (г. Ульяновск, 27–29 октября 2008 г.). В 2 т. Т. 2. Ульяновск: УлГТУ, 2008. С. 167–175.
8. Пащенко А.Е., Суворова А.В., Тулупьева Т.В., Тулупьев А.Л. Вероятностные распределения порядковых статистик в анализе сверхкоротких нечетких и неполных временных рядов // Труды СПИИРАН. 2009. Вып. 10. СПб.: Наука, 2009. С. 184–207.
9. Пащенко А.Е., Тулупьев А.Л., Суворова А.В., Тулупьева Т.В. Сравнение параметров угрожающего поведения в разных группах на основе неполных и неточных данных // Труды СПИИРАН. 2009. Вып. 9. СПб.: Наука, 2009. С. 252–261.

10. Пащенко А.Е. Математические модели и алгоритмы для комплекса программ по идентификации интенсивности рискованного поведения в условиях дефицита информации // Международная конференция по мягким вычислениям и измерениям. Сборник докладов. 2009. Т. 2. СПб.: Изд-во СПбГЭТУ «ЛЭТИ», 2009. С. 105–109.

11. Пащенко А.Е., Тулупьев А.Л., Тулупьева Т.В. Обработка нечеткости в ответах ВИЧ-инфицированных при оценке рискованности их поведения // Научная конференция МИФИ. 2007. Материалы. М.: МИФИ, 2007. Т. 3. С. 148–149.

12. Суворова А.В., Пащенко А.Е., Тулупьева Т.В., Тулупьев А.Л. Построение доверительных интервалов оценок интенсивности рискованного поведения на основе неравенства Чебышева // Труды СПИИРАН. 2009. Вып. 10. СПб.: Наука, 2009. С. 107–120.

13. Тулупьева Т.В., Тулупьев А.Л., Пащенко А.Е. Оценка интенсивности поведения респондента в условиях информационного дефицита // Труды СПИИРАН. Вып. 7. СПб.: Наука, 2008. С. 239–254.

14. Пащенко А.Е., Тулупьев А.Л., Николенко С.И. Моделирование заражения ВИЧ-инфекцией на основе данных о последних эпизодах рискованного поведения. // Известия высших учебных заведений: Приборостроение. 2006. №8. С. 33–34.

15. Тулупьева Т.В., Пащенко А.Е., Тулупьев А.Л., Красносельских Т.В., Казакова О.С. Модели ВИЧ-рискованного поведения в контексте психологической защиты и других адаптивных стилей. СПб.: Наука, 2008. 140 с.

16. Bell D.C, Trevino R.A. Modeling HIV Risk [Epidemiology] // JAIDS. 1999. Vol. 22(3). P. 280–287.

Abstract

The paper considers applied computer science problems related to estimating socially significant behavior parameters in case of initial data or knowledge granularity. The computational and methodological issues arise in intelligent systems that represent and process such data and knowledge are discussed. Several approaches to the issues elimination are proposed. The paper also introduces open research questions related to modeling and estimating the considered type of behavior.

Key words: information deficiency, data granularity, uncertain knowledge, behavior models, super-short imprecise series, risk estimates, rate estimates.

Суворова Алёна Владимировна, младший научный сотрудник лаборатории теоретических и междисциплинарных проблем информатики (ТuМПИ)

*Санкт-Петербургского института информатики и автоматизации РАН (СПИИРАН),
SUVALV@mail.ru,*

*Тулупьев Александр Львович, доктор физ.-мат. наук, доцент, заведующий лабораторией ТuМПИ СПИИРАН, доцент кафедры информатики математико-механического факультета СПбГУ,
ALT@ias.spb.su,*

*Пащенко Антон Евгеньевич, младший научный сотрудник лабораторией ТuМПИ СПИИРАН,
AEP@ias.spb.su,*

*Тулупьева Татьяна Валентиновна, канд. психол. наук, доцент; старший научный сотрудник лаборатории ТuМПИ СПИИРАН, доцент кафедры информатики математико-механического факультета СПбГУ, доцент кафедры психологии управления и педагогики СЗАГС,
TVT@ias.spb.su,*

*Красносельских Татьяна Валерьевна, канд. мед. наук, доцент кафедры дерматовенерологии с клиникой СПбГМУ им. акад. И.П. Павлова, начальник информационно-аналитического отдела Управления научных исследований СПбГМУ,
tatiana.krasnoselskikh@gmail.com*



Наши авторы, 2010.

Our authors, 2010.