

ПРИМЕНЕНИЕ ЭКСПЕРТНЫХ ОЦЕНОК В СИСТЕМАХ АВТОМАТИЗИРОВАННОЙ ДИАГНОСТИКИ¹

Р. М. Хвастунов, Ю. В. Новиков

Московский научно-исследовательский институт гигиены имени Ф. Ф. Эрисмана

Системы автоматизированной диагностики (АСД) и соответствующие центры в настоящее время создаются рядом медицинских вузов и научно-исследовательских институтов. Задача центра АСД заключается в проведении высококвалифицированной консультации по таким вопросам, как дифференциальная диагностика заболеваний, прогнозирование их возникновения и течения и назначение лечения. Алгоритмы вычислительной диагностики являются самым важным звеном АСД, так как от их качества зависит эффективность работы всей системы. Разработка алгоритмов вычислительной диагностики первоначально (в 60—70-х годах) осуществлялась с применением только математико-статистических методов, однако многолетний опыт показал, что точность заключений, осуществляемая этими алгоритмами, не только не превосходит точность работы опытных врачей [4, 6], но во многих случаях оказывается даже ниже [9].

Надежды на повышение точности вычислительной диагностики оказались связанными с применением методов экспертных оценок. При комбинировании экспертных и математико-статистических методов удалось разработать алгоритмы, обеспечивающие более высокую точность, чем у опытного врача [1, 5, 19], лимитированную лишь уровнем современной медицинской науки. Поэтому целесообразно обратить внимание специалистов, использующих АСД, на эффективность экспертных методов в разработке алгоритмов и одновременно таким образом осветить специфические методические трудности и нерешенные вопросы, чтобы в практической работе им было уделено заслуженное внимание.

Использование врачей как экспертов в связи с вычислительно-диагностическими и прогностическими алгоритмами имеет две стороны. Одни врачи выступают как эксперты при разработке алгоритмов. Ими являются, как правило, сотрудники НИИ и медвузов, которые хорошо подготовлены в области экспертных методов и не затрудняются в использовании формализованного языка, применяемого в АСД. Обеспечение качества экспертных оценок, производимых этими врачами, требует лишь достаточного времени, так как всегда можно провести повторный опрос, уточнить постановку вопросов и т. п. Другие же врачи — пользователи — предстают как эксперты при определении выраженности у больного тех или иных требуемых признаков, хотя даже минимально не знакомы с методами опроса экспертов и способами производства экспертных оценок.

Внедрение в практику АСД, естественно, привело к появлению систем дистанционной диагностики, то есть таких, когда врач-пользователь находится на значительном расстоянии от центра АСД и имеет с ним лишь телефонную связь. Поэтому разрабатываемые вычислительно-диагностические системы все более переходят на режим диалога с врачом-пользователем. Возрастает и будет возрастать роль его экспертных оценок, качество которых должно быть обеспечено на этапе разработки и апробации алгоритма, так как возможность повторного опроса и дополнительных разъяснений по постановке вопросов и т. п. в большинстве случаев отсутствует.

Применение экспертных методов при разработке алгоритмов имеет следующие специфические особенности:

1. Необходимость работы с малым числом экспертов (3—5 чел.), обычно опытных врачей, суждения которых могут быть использованы.
2. Возможность опроса этих врачей неоднократно и подолгу без ограничения формы опроса и времени.
3. При опросе применяются такие виды экспертных оценок, методы получения которых относительно мало разработаны, поэтому стоит задача разработки или усовершенствования методик опроса, позволяющих получить оценки этих видов.

Как работать с малым числом экспертов? Поскольку составляющие рабочую группу врачи-эксперты всегда являются представителями различных «школ» и их взгляды на диагностику и лечение, естественно, расходятся, то нельзя, по-видимому, рассчитывать получить от них всю относящуюся к делу информацию. Но она и не нужна. Разрабатываемый алгоритм всегда субъективен в том смысле, что как подбор необходимых признаков, так и способ операций с ними отражают знания и опыт имен-

¹ Статья написана по материалам доклада комиссии «Экспертные методы в разработке медицинских диагностических и лечебных поисковых систем» (председатель — проф. О. Б. Минскер) при Совете по кибернетике АН СССР и ВС НТО.

но той группы специалистов, которые участвуют в работе. Другая группа врачей-экспертов создала бы другой алгоритм. Поэтому задача организаторов заключается не в выявлении всего множества точек зрения (как это бывает, например, при проведении экспертизы для оценки качества продукции), а в том, чтобы наиболее адекватно отразить точку зрения имеющейся группы врачей-экспертов.

Из существующих методов опроса наибольшее использование при разработке алгоритмов вычислительной диагностики нашли интервью и анкетирование.

Интервью [3] проводится на первых этапах разработки для уточнения постановки задачи и формирования состава ориентирующей информации по заранее подготовленному списку вопросов, которые могут быть заданы эксперту. Однако не исключены значительные отступления от этого плана. В интервью обычно включают вопрос о характере и объеме документационной информации (историй болезни, других данных), на которую можно опереться. Желательно получить от врачей-экспертов сведения о возможности внедрения (где, в каком масштабе?) результатов проводимой разработки, что поможет лучше понять назначение и важность работы. В ходе интервью не следует стремиться получить от эксперта ответ на вопросы, в которых он чувствует себя неуверенно. Лучше способствовать максимальному развитию экспертом его мыслей по тем направлениям (в русле стоящей задачи), которые наиболее интересны и ясны самому эксперту. Целесообразно в ходе интервью выяснить также, в какой форме подготовить материал для дальнейшей работы эксперта (вид анкеты, состав ориентирующей информации).

Методом интервью удобно проводить группировку анализируемого материала (клинического архива историй болезни, вариантов принимаемых клинических решений и др.). Целесообразно при этом задавать вопросы о принципах группировки, логике соподчинения образуемых групп (если группировка иерархическая), что в дальнейшем поможет сократить поиски путей логического обоснования принимаемых клинических решений.

Анкетный опрос [3, 10] может быть эффективно применен для генерации совокупности признаков, необходимых для решения стоящей задачи, при постановке балльных оценок, разработке терминологии. Специфической особенностью анкетирования, осуществляемого для разработки алгоритмов вычислительной диагностики, является совершенно нетрадиционная форма анкеты. Анкета должна содержать что-то вроде предложений (в том числе и неверных) по решению предлагаемого экспертам вопроса. Например, можно поместить в анкету фрагмент схемы, приводящей к одному из клинических решений, и предложить эксперту построить аналогичные фрагменты к другим клиническим решениям.

Специально для построения «схем поиска диагноза» [12] предложен способ «мобильного анкетирования» [13]. При этом способе схема опроса меняется от эксперта к эксперту, что позволяет наиболее полно выявлять индивидуальную информацию эксперта. Повторные циклы интервью позволяют каждому эксперту ознакомиться с суждениями всех его коллег. Несмотря на имеющийся к настоящему моменту ряд методов эффективного опроса небольшого числа экспертов, поиски в этом направлении надо продолжать.

Вопросы получения и статистической обработки различных видов экспертных оценок. При решении задач клинической медицины применяют преимущественно следующие виды экспертных оценок: качественные, балльные, логические цепочки, группировки (классификации). Причем все эти виды оценок используют на этапе разработки алгоритма, на этапе же эксплуатации от врача-пользователя требуются только качественные оценки, производимые по признакам с заранее отработанными шкалами. Основные нерешенные и плохо разработанные вопросы методологии экспертных оценок связаны с их получением. Вопросы статистической обработки менее важны для практического применения.

Качественные экспертные оценки бывают двух типов. К первым относятся словесные описания, рисунки, фотографии каких-либо объектов: признаков и их градаций, состояний больного, типичной клинической картины заболевания и т. п. Почти всегда врачи для удобства переводят количественные показатели в качественную форму. Разделив шкалу количественного признака, например температуры, на несколько диапазонов с качественно определенным содержанием («норма», «субфебрилитет», «фебрилитет»), по существу получают шкалу качественного признака.

Качественные экспертные оценки второго типа производит врач-пользователь, когда он соотносит наблюдаемое им у больного проявление признака с тем или иным качественным описанием его градаций.

Первый вопрос, с которым приходится сталкиваться при использовании качественных оценок, связан с составлением шкал признаков. Плохо отработанное описание шкалы может привести к неверной оценке врачом-пользователем, который, может быть, иначе интерпретирует содержание терминов и понятий, используемых врачом-разработчиком.

Пример неудачного описания градаций: признак «Состояние сознания». Градации: 1 — ясное, 2 — помрачение, 3 — кома, 4 — эйфория.

Такое краткое описание может привести к тому, что, по нашим данным, лишь половина врачей, обследующих одного и того же больного, дадут совпадающие оценки состояния сознания. Необходимо составлять описания градаций так, чтобы даже

начинающие врачи давали совпадающие оценки [6, 19]. В приведенном примере удовлетворительным может быть следующее описание: 1) сознание ясное: больной активно ведет беседу с врачом, ориентирован в пространстве и времени; 2) помрачение сознания: больной отвечает на вопросы с запаздыванием по времени или даже не отвечает на часть вопросов, дезориентирован. Состояние оцепенения (ступор) и оглушения (сопор) необходимо отмечать как помрачение сознания; 3) кома: состояние, характеризующееся утратой сознания, отсутствием активных движений и реакции на болевые, световые и звуковые раздражения, нарушением вегетативных функций; 4) эйфория: состояние возбуждения, неадекватное общему состоянию больного и не оцениваемое критически самим больным. В связи с этим возникает вопрос: какова же методика обработки описаний и как убедиться в их достаточном качестве? Такая методика существует [17], хотя и для использования в технике, но она может быть применена и при решении задач вычислительной диагностики. Так, вначале составляется как можно больше различных описаний проявления интересующего признака или объекта. Описания берутся из литературы, подбираются разными врачами. Затем каждый врач по отдельности располагает эти описания в ряд, упорядоченный по принципу «лучше-хуже» или «больше-меньше». Из ранжированных рядов отбираются те описания, которые все врачи расположили одинаково относительно друг друга, и из данных описаний формируется требуемая шкала.

Для таких «интегральных» признаков, как «тяжесть состояния», «группа риска» и т. п., перед разработкой описаний градаций целесообразно выявлять те разные варианты клинических решений, которые будут приняты в зависимости от различия градаций. Опрос экспертов следует проводить не по вопросу упорядочивания описаний, а по вопросу: «Какие должны быть решения при данном описании градаций?».

Для отработки таких шкал удобно предоставлять экспертам в качестве контрольных примеров ряд условий, то есть не наблюдавшихся реально случаев заболеваний с тем или иным составом признаков, что позволит обойтись без подбора обширного клинического архива.

И все же применение этих способов не исключает ошибок со стороны врач-пользователя. Как специалист по узкому классу заболеваний он может плохо помнить или не знать специальной терминологии, относящейся к другим классам заболеваний. При разработке АСД, видимо, нужно предусматривать возможность запроса врачом-пользователем о способе выявления того или иного признака и его проявлений и создавать развернутые и подробные вспомогательные описания.

Количественной характеристикой качества шкал качественных признаков является воспроизводимость («надежность») [2, 18]. Эта величина Σ показывает вероятность ошибки 1 рода, то есть вероятность указания иной градации признака вместо фактически имеющейся у больного. Величина Σ может быть различной для разных градаций. Определить ее довольно трудно, так как нужно проводить опрос многочисленных врачей-экспертов. Опубликованные работы нам не известны, однако по косвенным данным Σ можно считать, что для плохо описанных градаций она составляет 0,20—0,50. В связи с этим стоит вопрос о вероятности ошибки в клиническом заключении, получаемом с использованием качественных показателей. Данный вопрос требует, по-видимому, дополнительных экспериментальных и теоретических исследований.

Балльные экспертные оценки использованы в большом количестве работ [8, 10, 11, 15, 19]. Однако далеко не во всех исследованиях достаточно подробно описан способ их получения, что, конечно, достойно сожаления, так как остаются, возможно, неизвестными найденные авторами оригинальные приемы.

Наиболее распространенным способом получения балльных оценок является использование вспомогательной шкалы [7, 10, 15]. Например: симптом патогномичен — 4 балла, очень характерен для диагноза — 3, характерен — 2, не характерен — 1, безразличен — 0.

В ходе опроса (анкетным способом) врач должен отнести каждую градацию того или иного признака к одной из градаций вспомогательной шкалы и присвоить ей соответствующий балл. Этот способ приводит обычно к хорошему совпадению оценок экспертов. Однако не бесспорны балльные оценки, использованные в самой вспомогательной шкале, так как неясен «принцип сопоставимости», то есть способ доказательства равенства одинаковых сумм балльных оценок, полученных при различных сочетаниях градаций. Можно ли, например, считать 4 симптома, не характерных для диагноза, равноценными одному патогномичному? Равноценны ли два нехарактерных симптома одному характерному?

Итак, использование вспомогательной шкалы обеспечивает лишь совпадение называемых оценок при совпадении суждений экспертов. Имеется несколько способов выполнения условия сопоставимости, однако они довольно жестко связаны с характером решаемой задачи, и их трудно применять для других задач. Например, при определении степени адаптированности ребенка к микросоциальным условиям детского учреждения сопоставимость балльных оценок обеспечивалась сравнениями комбинаций градаций. После предварительного назначения оценок по отдельным градациям признаков врачам-экспертам предъявлялись комбинации данных градаций, якобы имеющие место у сравниваемых детей, и врач должен был ответить, какое сочетание более благоприятно для адаптации или же какие комбинации в этом смысле равноценны. Например, 1-я комбинация: ребенок держится спокойно, не проявляя инте-

реса к контактам — «+1», засыпает с задержкой, без плача — «0»; 2-я комбинация: проявляет интерес к окружающим незнакомым лицам — «+3», сон беспокоен, с плачем — «-3».

Результат сравнения врача: первая комбинация показывает лучшую адаптированность ребенка. Поскольку сумма баллов первой комбинации больше, то балльные оценки не пересматриваются. В противном случае какая-то из балльных оценок была бы изменена [10].

Другой способ обеспечения условия сопоставимости — рассмотрение всех возможных сочетаний градаций признаков с подсчетом соответствующих сумм баллов [1]. Затем проверяется соответствие каждой из полученных сумм тому клиническому заключению, которое следует из врачебной логики для данного сочетания градаций. Этот полный перебор удобно начинать с минимальной (или максимальной) суммы баллов и затем постепенно продвигаться к границе, разделяющей различные заключения. Хотя эта процедура утомительна, но практически выполнима.

Важен еще один теоретический вопрос: постоянны ли балльные оценки значимости градаций некоторого признака при различных сочетаниях градаций других признаков? Представляется, что по крайней мере в некоторых случаях такое постоянство невозможно. Как же тогда установить факт непостоянства и как в подобном случае присваивать баллы? Удовлетворительное методическое решение данного вопроса пока неизвестно. Можно рекомендовать только метод полного перебора всех возможных сочетаний градаций, тогда и при непостоянстве баллов будет получено верное клиническое заключение, но этот путь не всегда может быть реализован. Поставленный вопрос еще ждет своей разработки.

Основные вопросы статистической обработки балльных оценок следующие: 1) определение согласованности (или несогласованности) оценок; 2) выбор целесообразного способа обобщения (целесообразной характеристики среднего) для согласованной группы оценок. Как же решать эти вопросы? Как известно, имеется ГОСТ 23554. 2—81 на обработку экспертных оценок, где в качестве критерия согласованности предлагается значение коэффициента вариации $C_v \leq 0,25$, однако теоретические или экспериментальные основания для этой рекомендации отсутствуют. По существу здесь применен эвристический прием, не учитывающий различия причин, обуславливающих расхождение экспертных оценок. Причинами расхождений оценок экспертов могут быть низкое качество вспомогательной шкалы либо принципиальные расхождения суждений врачей о клинической значимости признаков, либо еще что-нибудь. В разных случаях надо применять различные критерии согласованности и разные меры повышения этой согласованности. Нельзя механически отбрасывать «выскакивающие» оценки отдельных врачей. Врач-эксперт, как правило, не забывает свои оценки и не может работать в условиях, когда ему неясны принципиальные причины расхождения с другими врачами-экспертами. Таким образом, вместе с выброшенной оценкой имеется риск потерять и врача-эксперта или его доверие к методам экспертных оценок, что одинаково нежелательно.

В качестве обобщающей характеристики оценок согласованной группы также нельзя без размышлений принимать среднее арифметическое. Зачастую целесообразно применение медианы или моды, а в ряде случаев следует выбирать минимальную или же максимальную из предложенных экспертами оценок. Например, при прогнозировании исходов оперативного лечения сердечно-сосудистых заболеваний рационально указывать наименее благоприятный исход, а поэтому обобщающей характеристикой оценок, присвоенных экспертами показателю тяжести состояния, должна быть наибольшая из индивидуальных оценок.

Разработка рекомендаций по статистической обработке балльных оценок в приложении к задачам вычислительной диагностики представляет, по-видимому, давно назревшую задачу.

Логические цепочки. При разработке вычислительных алгоритмов в задачах клинической медицины часто пользуются экспертными оценками, имеющими вид логических условий, когда в зависимости от значений тех или иных признаков меняется способ дальнейших действий. Задача получения от экспертов серий таких действий, то есть логических цепочек, вероятно, впервые сформулирована О. Б. Минскером и Л. Ф. Ведмеденко. С использованием логических цепочек оказалось возможным строить алгоритмы типа «схем поиска диагноза», которые оказались высокоэффективными, что повысило интерес к этому виду экспертных оценок.

Сейчас имеется уже несколько способов получения от врачей-экспертов логических цепочек, которые приводят к решению соответствующей задачи вычислительной диагностики. Например, способ «Первый эксперт — группа» (название наше) состоит в следующем. Врач-составитель, выступающий в качестве первого эксперта, выбирает наиболее значимый, по его мнению, признак (или признак, с которого принято начинать исследование). Врач должен ответить на вопрос: можно ли дать какое-либо клиническое заключение или рекомендацию на основе отдельных градаций упомянутого признака? В случае утвердительного ответа на этот вопрос к соответствующей градации присоединяется с помощью стрелки блок, содержащий данное заключение. Если же рекомендация не может быть выдана, то врач должен указать на следующий признак, исследование которого нужно для выдачи рекомендаций. В результате появляется следующее звено алгоритма, по отношению к которому вопрос повторяет-

ся. Путем перебора всех градаций каждого признака удается построить логическую систему выдачи заключений. В отдельных звеньях ее возможно введение вычислительных операций, и тогда получается логико-вычислительный алгоритм.

Задача второго и последующих экспертов — проверка обоснованности рекомендаций первого эксперта, отраженных в схеме. При этом врачи могут вводить новые признаки, новые логические ходы, изменять содержание клинических заключений. После проверки схемы несколькими врачами она возвращается к первому эксперту, который выступает уже на общих основаниях.

Надо отметить, что техника опроса экспертов при построении логических цепочек пока еще несовершенна, громоздка и плохо формализована. Нужны работы по ее дальнейшей формализации с учетом следующих вопросов: 1) как при опросе гарантировать полноту охвата вариантов клинического состояния больного, требующих решений? 2) как строить дублирующие варианты логики (то есть как подойти к одному и тому же клиническому решению разными путями)? 3) какова степень надежности получаемого заключения, если градации признаков могут быть определены с ошибкой?

Математико-статистической проработки требуют следующие вопросы теории логических цепочек.

1. Определение качества предложенных логических цепочек, то есть необходимо выяснить, по каким показателям сравнивать между собой предложенные различными врачами варианты логических цепочек, ведущих к одному и тому же набору клинических заключений. Видимо, здесь может быть одна группа показателей, характеризующих легкость использования цепочки (ее краткость?) и другая — результаты практического использования в алгоритме (надежность, вероятность ошибки, подробность).

2. Вопросы обобщения (объединения, усреднения?) различных цепочек, направленных на выработку какого-то круга решений. Для этого, видимо, надо ввести меру близости различных цепочек, найти характеристику рассеяния для их совокупности.

Таким образом, наблюдающаяся тенденция расширения круга задач клинической медицины, решаемых с применением математико-статистических и экспертных методов, а также развитие АСД в диалоговом режиме повышают важность решения методических вопросов экспертных оценок. Основными из них представляются следующие:

1. Дальнейшая разработка методов опроса малого числа экспертов, в том числе способов организации дискуссии между ними при различных видах производимых экспертных оценок.

2. Дальнейшее совершенствование способов описания градаций качественных признаков с целью составления таких описаний, которые полностью исключили бы возможность ошибки со стороны врача-пользователя.

3. Совершенствование методов получения всех видов экспертных оценок, в первую очередь балльных оценок и логических цепочек, как играющих наибольшую роль в алгоритмах вычислительной диагностики.

4. Разработка методов статистической обработки различных видов экспертных оценок с учетом метода их получения и характера решаемой с их помощью задачи.

Поступила 03.12.85.

ОБМЕН ОПЫТОМ И АННОТАЦИИ

УДК 616—053.2—084.3

Т. М. Шигабудинова (г. Нижнекамск, ТАССР). Опыт диспансеризации детей на педиатрическом участке

На педиатрическом участке состоят на учете 840 детей, из них в возрасте до года — 66, «неорганизованных» — 140. В ходе проведения диспансеризации было изучено состояние здоровья детей по данным профилактических осмотров, а также материалы участкового терапевта о состоянии здоровья семьи.

При оценке состояния здоровья детей использованы критерии, приведенные в методических рекомендациях Минздрава СССР «Комплексная оценка состояния здоровья детей и подростков при массовых врачебных осмотрах» от 27.04.82 г. (см. табл. 1).

Для оздоровления детей используются все возможные лечебно-профилактические учреждения (см. табл. 2). Оздоровление детей II потока школьного возраста в стационаре не всегда осуществимо. Кроме того, при госпитализации школьники отстают от школьной программы. Поэтому при детской поликлинике организовано отделение реабилитации, где проводятся дополнительные исследования и осмотр врачами узких специальностей. Производятся лечение очагов хронической инфекции, ЛФК, массаж, психо- и рефлексотерапия. Подбирается правильный режим питания и физической нагрузки. В большинстве случаев исключается нерациональное и не всегда оправданное