

А.А.Болотов, В.Г.Барчуков

МЕТОДЫ И ПРИНЦИПЫ ПОСТРОЕНИЯ ПОЛНЫХ И НЕПРОТИВОРЕЧИВЫХ БАЗ ЗНАНИЙ ПРИ ДИФФЕРЕНЦИАЛЬНОЙ ДИАГНОСТИКЕ ДВУХ БЛИЗКИХ ПО СИМПТОМАТИКЕ ЗАБОЛЕВАНИЙ

ГНЦ ФМБЦ им. А.И. Бурназяна ФМБА России, Москва

Контактное лицо: Болотов Александр Александрович: ABolotov @bk.ru

Резюме

Предмет, тема, цель. В статье рассматриваются методы, принципы и подходы к построению полных и непротиворечивых баз экспертных знаний в задачах дифференциальной классификации, основанные на психологически корректных способах получения информации от эксперта.

Методология проведения. Рассмотрены проблемы передачи знаний от эксперта компьютерной программе. Важнейшими характеристиками процедуры извлечения знаний является ее психологическая корректность, полнота извлечения и представления знаний, а также непротиворечивость (безошибочность) базы знаний. Поскольку основным источником информации является человек, то для соблюдения этого правила, обязательным условием должно быть использование его логики мышления.

Результаты. Предложены рациональные методы и алгоритмы по формированию больших баз знаний для дифференциальной диагностики двух близких по симптоматике заболеваний на основе психологически корректного способа опроса эксперта, решающего привычную для него задачу диагностики по полному описанию случая в текстовом виде. Приводятся результаты статистического моделирования методом Монте-Карло процесса извлечения экспертных знаний для таких задач в рамках алгоритма, основанного на поиске границ, разделяющих два заболевания по схеме двоичного приближения от так называемого среднего объекта, имеющего половину признаков наиболее характерных для одного заболевания и половину признаков наиболее характерных для другого заболевания при различном количестве используемых признаков и тремя видами возможных границ между заболеваниями.

Область применения результатов. Результаты работы имеют прикладное значение для построения автоматизированных рабочих мест практических врачей с целью повышения качества диагностических заключений и оптимизации лечения пациентов.

Выводы. Психологически корректные принципы извлечения знаний, методы и алгоритмы, изложенные в статье, позволяют быстро и надежно строить полные и непротиворечивые базы знаний большого объема в различных областях дифференциальной диагностики заболеваний для создания соответствующих автоматизированных рабочих мест врача, что позволит улучшить качество принимаемых врачебных решений, особенно в сложных случаях.

Ключевые слова: полные и непротиворечивые базы знаний, дифференциальная диагностика, психологическая корректность извлечения знаний, алгоритм двоичного поиска границ между заболеваниями, метод Монте-Карло

Для цитирования: Болотов А.А., Барчуков В.Г. Методы и принципы построения полных и непротиворечивых баз знаний при дифференциальной диагностике двух близких по симптоматике заболеваний // Клинический вестник ФМБЦ им. А.И. Бурназяна 2022. № 4. С. 60–66. DOI:10.33266/2782-6430-2022-4-60-66

DOI:10.33266/2782-6430-2022-4-60-66

A.A.Bolotov, V.G.Barchukov

Methods and Principles of Building Complete and Consistent Knowledge Bases in the Differential Diagnosis of two Diseases Similar in Symptoms

A.I. Burnasyan Federal Medical Biophysical Center, Moscow, Russia

Contact person: Bolotov Alexandr Aleksandrovich: iABolotov @bk.ru

Abstract

Subject, topic, purpose. The article discusses methods, principles and approaches to the construction of complete and consistent expert knowledge bases in differential classification problems based on psychologically correct ways of obtaining information from an expert.

Methodology. The problems of transferring knowledge from an expert to a computer program are considered. The most important characteristics of the knowledge extraction procedure are its psychological correctness, the completeness of the extraction and presentation of knowledge, as well as the consistency (infallibility) of the knowledge base. Since the main source of information is a person, in order to comply with this rule, the use of his logic of thinking must be a prerequisite.

Results of the work. Rational methods and algorithms for the formation of large knowledge bases for the differential diagnosis of two diseases similar in symptoms are proposed on the basis of a psychologically correct method of interviewing an expert who solves the usual diagnostic task for him by a complete description of the case in text form. The results of Monte Carlo statistical modeling of the process of extracting expert knowledge for such tasks are presented within the framework of an algorithm based on the search for boundaries separating two diseases according to the binary approximation scheme from the so-called average object having half of the signs most characteristic of one disease and half of the signs most characteristic of another disease with a different number of used signs and three types of possible boundaries between diseases.

Scope of application of the results. The results of the work are of applied importance for the construction of automated workplaces of practitioners in order to improve the quality of diagnostic conclusions and optimize the treatment of patients.

Conclusions. Psychologically correct principles of knowledge extraction, methods and algorithms described in the article allow you to quickly and reliably build complete and consistent knowledge bases of large volume in various areas of differential diagnosis of diseases to create appropriate automated doctor's workplaces, which will improve the quality of medical decisions, especially in difficult cases.

Keywords: complete and consistent knowledge bases, differential diagnosis, psychological correctness of knowledge extraction, binary search algorithm for boundaries between diseases, Monte Carlo method

For citation: Bolotov AA, Barchukov VG. Methods and Principles of Building Complete and Consistent Knowledge Bases in the Differential Diagnosis of two Diseases Similar in Symptoms. A.I. Burnasyan Federal Medical Biophysical Center Clinical Bulletin. 2022.4:60-66. (In Russian) DOI:10.33266/2782-6430-2022-4-60-66

Введение

Повышение уровня качества и эффективности оказания медицинской помощи является одной из актуальных задач современной медицины. Однако удаленность медико-санитарных частей ФМБА России от ведущих лечебных учреждений страны не редко затрудняет в сложных клинических ситуациях постановку точного диагноза, что определяет поиск методов, позволяющих практическому врачу, правильно поставить диагноз и назначить наиболее оптимальное лечение. Особенно это важно, когда возможности консультации с ведущими специалистами страны крайне ограничены.

Одним из возможных путей улучшения качества диагностики в настоящее время рассматривают создание компьютерных систем, основанных на знаниях высококвалифицированных врачей и создания соответствующих автоматизированных рабочих мест (далее – АРМ) [1].

Однако на пути создания таких АРМов стоит проблема передачи знаний от врача-эксперта компьютерной программе, известная как проблема приобретения знаний [1]. В основном здесь речь идет о так называемых знаниях 2-го рода или умениях [1,2]. Такие знания обычно не входят в учебники и справочники. Они представляют категорию личностных знаний, интуицию и искусство решения проблем, а это лежит в области подсознательного [1,3]. И как следствие из подсознательного характера экспертных знаний следует невозможность их выявления путем прямого опроса экспертов. Однако эксперты могут сообщить факты, передать декларативное знание, но при этом они неспособны вербализовать свои процедуральные знания - знания практического типа т.е. правила принятия диагностических заключений.

Имеющиеся литературные данные [4,5,6] свидетельствуют о том, что принятие решений происходит в кратковременной (рабочей) памяти человека. Объем кратковременной памяти ограничен «магическим» числом 7 ± 2 чанка (чанк - фрагмент знаний, хранимый и используемый как единое целое). Ограничение объема кратковременной памяти оказывает существенное влияние на систему переработки информации человеком и именно это ограничение заставляет людей вырабатывать чанки все более емкого и обобщенного характера, т.е. ограниченный объем кратковременной памяти заставляет экспертов приспосабливать классификационные задачи к своим возможностям. Прежде всего, в случае нескольких классов решений эксперты заменяют параллельную задачу (классификация объектов на несколько классов) на

иерархическую совокупность последовательных задач: деление на две группы, затем на подгруппы и уже потом - на два класса. Другими словами, при дифференциальной диагностике врач интуитивно строит иерархическую систему симптомов заболевания, начиная с наиболее общего синдрома комплекса, и доходит до конкретного симптома комплекса (заболевания). При этом он постоянно осуществляет разделение на два синдрома комплекса, выбирая наиболее характерный, в который вероятнее всего входит конкретный случай (заболевания). И это разделение продолжается до принятия врачом окончательного диагностического решения.

К таким результатам приводят многочисленные эксперименты по изучению принятия решений человеком [4,5]. Эти работы показали, что использование иерархических схем является одним из приемов, позволяющим избежать чрезмерной нагрузки на кратковременную память при воспроизведении информации и принятии решений [6,7].

Таким образом, обобщая вышеизложенное, можно заключить, что вопрос о выявлении и передаче компьютеру человеческих знаний и умений является крайне сложным. На пути решения этой задачи стоит множество проблем.

Цель

С учетом вышеизложенного, целью данной статьи является разработка подходов и методов психологически корректных способов извлечения знаний от экспертов для задач дифференциальной диагностики заболеваний и создания компьютерных систем, содержащих полные и непротиворечивые базы знаний, копирующих в полном объеме суждения эксперта.

Методы исследования

Важнейшей характеристикой процедуры извлечения знаний является ее психологическая корректность. Поскольку основным источником информации является человек, то для соблюдения этого правила, обязательным условием должно быть использование его логики мышления. При этом известно [1,2], что количество психологически корректных операций получения информации в многоаспектных задачах сравнительно невелико. Но только эти способы ведут к надежным базам знаний большого объема.

Многочисленными психологическими экспериментами показано, что люди ведут себя надежно и непротиворечиво при ограниченном числе возможных путей (классов) решения, числе симптомов

(или признаков) и числе возможных значений для каждого из симптомов [1,5,6]. Эти ограничения являются определяющими при разработке методов извлечения экспертных знаний.

Еще одной важной характеристикой, связанной с получением информации от экспертов, является полнота извлечения и представления знаний, а также непротиворечивость (безошибочность) базы знаний. Под полнотой и непротиворечивостью базы знаний, в общем смысле, понимается способность системы формировать непротиворечивое решение для любой ситуации в рамках ограниченной предметной области.

Среди источников неполноты и ошибочности базы знаний (БЗ) могут быть выделены следующие два фактора:

1. Знания и умения эксперта на когнитивном уровне, отражающие его квалификацию и степень проработанности проблемной области, могут быть неполными и содержать ошибочные представления о реальности;

2. При передаче знаний компьютеру могут появляться ошибки, связанные с неточностью и неоднозначностью естественного профессионального языка эксперта, так и с вербализацией тех или иных представлений эксперта. Кроме того, с одной стороны эксперт может и не понимать, какие именно из его знаний необходимо эксплицировать, а с другой стороны не все знания эксперта могут быть адекватно вербализованы. При отображении знаний с лингвистического на машинный уровень представления знаний также возможно уменьшение точности и полноты описания реальности.

Следовательно, чтобы построить базу знаний надежным образом, в ходе ее построения должны быть использованы лишь вопросы к экспертам, допустимые с точки зрения человеческой системы переработки информации, любые ответы экспертов проверяются на непротиворечивость, вопросы к экспертам представлялись в формате имитации обычной работы, выполняемой ими в сфере их профессиональной деятельности.

Собственные исследования

Важно отметить, что только полная и непротиворечивая база знаний, полученная на основе полной совокупности граничных объектов и описывающих их решающих правил позволяет выносить наиболее правильное решение. Совокупный учет отмеченных фактов и соблюдение указанных ограничений позволил создать программный комплекс, позволяющий психологически корректно и высоко профессионально (с высокой долей вероятности) помогать принимать врачу правильные диагностические решения и наиболее адекватную состоянию больного лечебную тактику. При этом для формирования, хранения, анализа, манипулирования и использования полных и непротиворечивых баз знаний оказалось целесообразным использовать подходы на основе систем управления базами данных (СУБД) реляционного типа, хорошо подходящих для таких целей и имеющих мощные средства в виде языка

SQL (структурный язык запросов) для поиска и многостороннего анализа баз знаний.

При разработке системы извлечения экспертных знаний мы использовали следующие подходы:

1. Подсознательный характер экспертных знаний заставил нас отказаться от попыток получить от эксперта решающие правила и выбрать как основную стратегию извлечения знаний - предъявление эксперту описаний объектов.

2. Полная база экспертных знаний, позволяющая сделать экспертную систему действительно «знающей», готовой дать ответ на любой вопрос и при этом быть уверенным в качестве её работы, определила нашу направленность на построение так называемых полных и непротиворечивых баз медицинских знаний.

3. Ориентация на возможность создания больших баз знаний, содержащих десятки тысяч продукций, определила необходимость введения ряда ограничений на структуру признаков, в частности:

а) использованы могут быть только дискретные признаки в формате: признак-градация;

б) число признаков и их градаций должно быть по возможности минимальным;

в) целесообразно использовать только те градации признаков, которые наиболее значимы для дифференциальной диагностики.

Не требует доказательства и является очевидным, что полная база знаний - это результат экспертного разделения возможных вариантов рассматриваемых заболеваний.

Полная база знаний содержит набор продукций, т.е. множества объектов, заданных вектором признаков (т.е. описание случая в виде линейной комбинации градаций используемых признаков) и классом (т.е. заболеванием), к которому экспертом отнесен этот случай.

При этом основной вес в разделении рассматриваемых заболеваний принадлежит граничным объектам, для которых изменение значения только одного диагностического признака влечет за собой переход в другой класс решений, т.е. свидетельствует о наличии другого заболевания. Множество граничных объектов определяет границу, разделяющую классы решений. Следовательно, если определить границу между рассматриваемыми заболеваниями, то задача диагностики будет решена.

Как уже отмечено выше полные базы экспертных знаний, включающие тысячи и десятки тысяч продукций, практически нельзя получить от эксперта последовательным предъявлением их всех. При этом одним из путей решения проблемы построения таких баз знаний является поиск граничных вариантов заболеваний с исключением наиболее характерных вариантов рассматриваемых заболеваний. Для решения этой проблемы мы использовали известный принцип «доминирования по характерности» [1,2]. В соответствии с этим принципом решение эксперта об отнесении конкретного случая к определенному заболеванию (классу А) означает, что к этому же классу относится и подмно-

жество симптомокомплексов (случаев) более характерных для этого заболевания. Они автоматически включаются в информационное поле возможных его вариантов и в последующем эксперту не предъявляются, что позволяет существенно сократить время опроса. Иначе говоря, к классу А относятся все объекты, доминирующие по «типичности» данный случай (объект базы знаний).

При построении полных баз знаний не редко возникают ситуации, когда врач-эксперт дает противоречивые заключения, которые необходимо выявлять и после проведения дополнительного анализа выбирать, наиболее правильный.

Поиск противоречий в заключениях эксперта основан на использовании уже приведенного выше принципа доминирования по характерности. Другими словами, если имеется два рядом стоящих варианта описания случая заболевания и один отнесен экспертом к одному заболеванию, а второй к другому, но при этом значения симптомов второго случая более характерно для первого заболевания, то возникает противоречие, которое и устраняется повторным предъявлением его врачу-эксперту.

Результаты

Для решения задач дифференциальной классификации между двумя близкими по симптоматике заболеваниями, нами была разработана система МКЛАСС с использованием подходов системы ДИФКЛАСС [8], в основу которой положен принцип поиска граничных вариантов заболеваний на основе двоичного приближения от среднего объекта.

Средний объект представляет собой вариант сочетания симптомов половина, из которых характерна для одного заболевания, а другая половина для второго заболевания.

В общем виде задачу построения полных и непротиворечивых баз знаний для дифференциальной диагностики можно представить следующим образом. Имеется два близких по симптомокомплексу заболевания А и В, которые в свою очередь имеют N число симптомов (признаков) с двумя градациями. Одна градация, например, первая – всегда наиболее характерна для заболевания А, другая – вторая – для В. При рассмотрении возможных вариантов может быть несколько решений:

- либо имеет место заболевание А;
- либо имеет место заболевание В;
- либо имеет место случай, когда у пациента есть и заболевание А, и заболевание В вместе, т.е. случай С;
- либо имеет место случай, когда у пациента отсутствует как заболевание А, так и заболевание В, т.е. случай D.

Для случаев С и D значения каждого из отдельных признаков равно характерны, т.е. рассматриваемые признаки не типичны для этих случаев, и принцип доминирования по характерности для них не используется.

Таким образом, для построения полной базы знаний в задаче дифференциальной диагностики заболеваний с N признаками, имеющих по две гра-

дации в общем случае эксперту необходимо представить 2^N случаев для классификации. При использовании 12 признаков для диагностики это будет составлять 4096 случаев. Для оптимального опроса эксперта и ускорения построения полных и непротиворечивых баз знаний ему необходимо представлять граничные объекты, кроме того, целесообразно использовать любые дополнительные сведения о структуре задачи для ускорения заполнения базы знаний.

В системе МКЛАСС [9] были расширены возможности создания полных и непротиворечивых баз знаний большого объема за счет экспертного определения несовместных признаков, используемых для дифференциальной диагностики. В этом случае все комбинации признаков с указанными несовместными признаками будут в полной базе решающих правил (т.е. в базе знаний) отнесены к классу D. Так при 12 бинарных признаках задание несовместности для двух признаков позволяет определить 1024 случая из полной базы знаний содержащей 4096 продуктов, таким образом для опроса эксперта требуется предъявить всего 3072 случая.

Также в МКЛАСС для ускорения построения полных и непротиворечивых баз знаний была реализована методика выявления эвристических правил эксперта на основе последовательного предъявления средних объектов для классификации. При последовательном предъявлении средних объектов каждый объект от другого отличается значениями всего двух признаков. Это позволяет продвигаться по среднему слою (совокупности средних объектов) с минимальными изменениями в признаках, поэтому при наличии каких-либо наличия каких-либо устойчивых причинно-следственных отношений признаками, как оказалось на практике, они достаточно легко подмечаются экспертом. Попытки же «выспросить» подобные правила у эксперта перед началом опроса были, как-правило, мало результативны. Обычно эксперту удавалось вербализовать 3-4 правила, которые отражали, в основном, типичные причинно-следственные отношения, присущие рассматриваемой задаче, и которые не давали заметного ускорения заполнения базы знаний, поскольку они не могли осуществлять расширение базы знаний по доминированию.

Рассмотрим более подробно этот процесс. Допустим, что эксперту предъявляется следующая последовательность средних объектов, заданная векторами признаков для 6-ти бинарных признаков (при этом предполагается, что первые оценки признаков наиболее характерны для первого класса, а вторые – для второго класса):

- 1) 1 2 1 1 2 2
- 2) 1 2 1 2 1 2
- 3) 1 2 1 2 2 1

Допустим, что эти три объекта эксперт отнес к 1-му классу решений. Из примера видно, что другим общим свойством этих объектов является наличие комбинации первых 3-х признаков: »1 2 1». Значит подобная комбинация признаков опреде-

ляет вполне определенную эвристику. А т.к. второй признак в этих цепочках менее характерен для 1-го класса решений, то и комбинация первых трех признаков «1 1 1» также приведет к заключению, что объекты, имеющие эти признаки будут относиться к 1-му классу. Тогда можно сформулировать следующие правила отнесения ряда объектов к 1-му классу:

1 * 1 1 * *
1 * 1 * 1 * (1)
1 * 1 * * 1

т.к. первые оценки последних 3-х признаков еще более характерны для 1-го класса. Символом «*» обозначен факт возможного наличия любой оценки того или иного признака. Из этих трех правил можно сформировать одно, т.к. последние 3 признака не оказывают влияния на отнесение таких объектов к 1-му классу:

1 * 1 * * * (2)

Задавая три правила вида (1) или одно правило вида (2) мы определяем класс для всех объектов, удовлетворяющих этим условиям.

В качестве основного метода для эффективного построения полных и непротиворечивых баз знаний в задачах дифференциальной диагностики используется алгоритм двоичного приближения к граничным вариантам заболеваний от так называемого среднего объекта. Суть его заключается в следующем:

Алгоритм автоматически формирует средний объект. Как отмечено выше средний объект (V_{cp}), это случай заболевания, когда половина симптомов (признаков) имеет градацию 2, а половина - 1. Такой случай заболевания предъявляется эксперту.

Пусть рассматриваемый средний объект (V_{cp}) по решению эксперта принадлежит классу (заболеванию) А. При этом граничный объект будет расположен в N-мерном пространстве, где-то между V_{cp} и вектором самого типичного случая для заболевания В (V_b), (т.е. случай когда весь набор признаков имеет градацию 2). Далее все компоненты V_{cp} , имеющие градацию 2, будут неизменными, а заменяться градацией 2 будет половина признаков, имеющих градацию 1. Возможно два варианта: первый, когда заменяемое число признаков, имеющих градацию 1 будет четным и второй, когда заменяемое число признаков имеющих градацию 1 будет нечетным. В первом случае такое число признаков будет равно $N/4$, а во - втором, либо $0.25(N-1)$, либо $0.25(N+1)$.

Таким образом, сформировались два новых варианта описания заболевания (два вектора V_1 и V_2). Далее векторы V_1 и V_2 предъявляются эксперту. Если какой-то из этих векторов принадлежит классу А, то описанная процедура повторяется снова. Такое деление идет до тех пор, пока все градации признаков 1 заболевания А не будут переведены в градацию 2. После каждого ответа эксперта происходит автоматическое заполнение базы знаний с использованием принципа доминирования по характерности.

Очевидно, что если средний объект по решению эксперта принадлежит классу (заболеванию) В, то алгоритм построения будет аналогичен, но изменяемой частью являются признаки с градацией 2.

В случае, когда средний объект по решению эксперта принадлежит одному из классов С или D, то алгоритм поиска граничного случая заключается в том, что за средний объект условно берется сначала заболевание А и от него, как описано выше, изменяя градации 1 на 2 формируются новые варианты, которые представляются эксперту для принятия решения об отнесении его к одному из рассматриваемых (А, В, С или D) классам (заболеваниям). Затем условно берется заболевание В и алгоритм повторяется. Иными словами, из объекта, принадлежащего дополнительному классу, строятся цепочки «вверх» и «вниз». Объекты, отнесенные экспертом к дополнительным классам, просто заносятся в базу знаний без распространения по доминированию.

Таким образом, алгоритм позволяет найти все граничные объекты.

Однако остается открытым вопрос об универсальности этого алгоритма и скорости заполнения такой базы знаний в зависимости от типа границы (расположение в N мерном пространстве признаков, рассматриваемых заболеваний), а также обеспечения ее полноты. Для решения этой задачи нами был использован общеизвестный принцип статистического моделирования (метод Монте-Карло). В качестве тестового объекта исследования взяли 3 типа границ, которые по имеющимся данным [4,10] могут составлять реальные границы между классами решений (заболеваний) в реальных задачах дифференциальной диагностики:

1) простая граница (типа «отсечка») – тип границы, когда случайным образом берется признак и по номеру его градации фиксируется класс (заболевание), т.е. все случаи с данной градацией признака будут иметь этот же класс, а все остальные - другой;

2) случайная граница, когда каждый случайный выбранный вариант заболевания случайным образом относится к определенному классу решения (заболевания), при этом в процессе построения непротиворечивой базы знаний обязательным является использование принципа распространения по доминированию;

3) детерминированная (линейная) граница – граница, построенная на базе средних объектов всех рассматриваемых признаков дифференцируемых заболеваний. В теоретическом смысле это самая длинная и сложная граница.

Были рассмотрены случаи дифференциальной диагностики с четырьмя, шестью и восемью бинарными признаками.

Результаты такого моделирования приведены на рис. 1-3.

При этом по оси ординат отложено относительное заполнение базы знаний (в %), а по оси абсцисс - число объектов (вариантов описания заболевания), которые гипотетически могли быть предложены эксперту для классификации:

граница №1 – верхний график (красный цвет);
граница №2- средний график (зеленый цвет);
граница №3 – нижний график (синий цвет).

Из рисунка видно, что для всех типов границ и числа рассматриваемых признаков алгоритм позво-

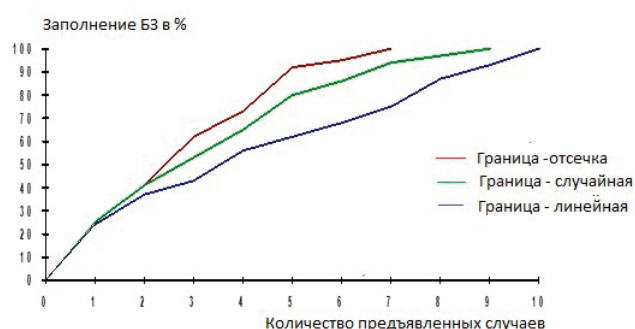


Рис. 1. Результаты моделирования заполнения базы знаний для 3-х границ при 4 признаках (полная база знаний содержит 32 продукции)
Fig. 1. The results of modeling the filling of the knowledge base for 3 boundaries with 4 signs (the complete knowledge base contains 32 products)

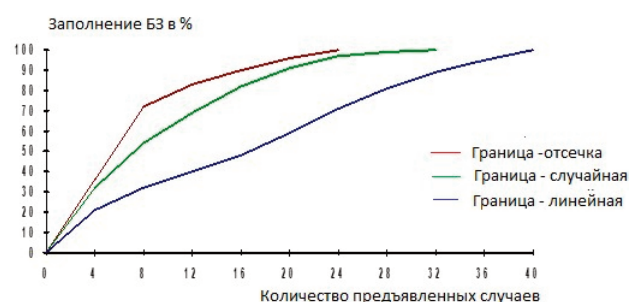


Рис. 2. Результаты моделирования заполнения базы знаний для 3-х границ при 6 признаках (полная база знаний содержит 64 продукции)
Fig. 2. The results of modeling the filling of the knowledge base for 3 boundaries with 6 signs (the complete knowledge base contains 64 products)

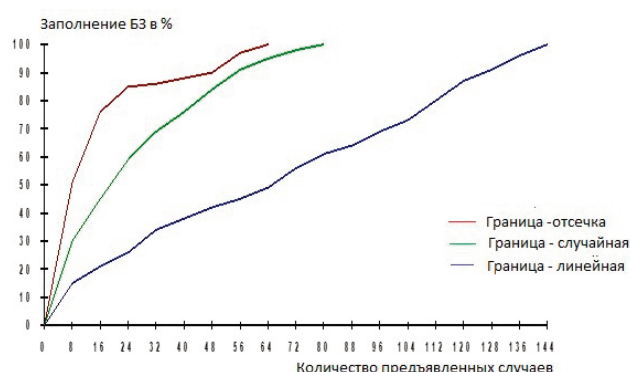


Рис. 3. Результаты моделирования заполнения базы знаний для 3-х границ при 8 признаках (полная база знаний содержит 256 продукции)
Fig. 3. The results of modeling the filling of the knowledge base for 3 boundaries with 8 signs (the complete knowledge base contains 256 products)

ляет строить базы знаний в полном объеме (100%). Однако в зависимости от типа границы число итераций (число возможно необходимых предъявлений эксперту) различно. Для границы типа «отсечка» оно минимально, тогда как для границы, построенной на базе средних объектов, она максимальна. Промежуточное положение занимает граница 2 – случайная граница. Однако именно такая граница наиболее вероятна что в реальных задачах. Следовательно результаты моделирования будут близки к реальным оценкам реальных задач. Т.е. в реальных условиях заполнение базы знаний будет достаточно быстрым. Следует заметить, что заполнение БЗ на начальных этапах опроса идет более быстрыми темпами, чем на завершающих для границ типа 1 и 2 (заполнение 50% базы знаний требует 20 до 30% итераций). При этом видно, что эффективность алгоритма увеличивается при увеличении числа признаков, т.е. отношение числа объектов, представленных эксперту к общему числу объектов в базе знаний при увеличении числа используемых признаков, уменьшается. Если учесть, что в реальных условиях используют от 10 до 15 признаков, то скорость заполнения БЗ будет достаточно высокой.

На основе вышеизложенных подходов было построено более 5 реальных баз знаний по дифференциальной диагностике тромбоэмболии легочной артерии с инфарктом миокарда, острой крупозной пневмонией, острым животом и т.п. Среднее время работы (интервьюирования) эксперта по формированию каждой из таких баз знаний для дифференциальной диагностики при использовании 12 признаков составило около 2 часов. Построение этих баз знаний подтвердило универсальность предложенных подходов, отсутствие у них зависимости от типа границы, а также высокую скорость заполнения базы экспертных знаний.

Выводы

1. Психологически корректные принципы извлечения знаний, заложенные в системе МКЛАСС позволяют быстро и надежно строить полные и непротиворечивые базы знаний.

2. Алгоритм двоичного приближения к граничным вариантам заболеваний от средних объектов (вариантов заболеваний, имеющих половину признаков наиболее типичных для одного заболевания, а половину – другого заболевания) позволяет эффективно строить полные и непротиворечивые базы знаний для любых границ между рассматриваемыми заболеваниями.

3. При использовании предложенных подходов в системе МКЛАСС появляется возможность строить большие полные и непротиворечивые базы знаний для дифференциальной диагностики заболеваний, содержащих десятки тысяч продукции за короткое время опроса эксперта.

СПИСОК ИСТОЧНИКОВ

1. Ларичев О.И., Мечитов А.И., Мошкovich Е.М., Фуремс Е.М. Выявление экспертных знаний. М.: Наука, 1989. 128 с.
2. Ларичев О. И. Теория подсознательных решающих правил и ее применение в диагностических задачах // Психологический журнал. 2003. Т. 24, № 1. С. 56–64.
3. Kilstrom J. The Cognitive Unconscious // Science. 1987. V.237, No. 4821. P. 1445-1452.
4. Ларичев О.И. Структуры экспертных знаний в задачах классификации // Доклады Академии наук. 1994. Т.336, № 6, С. 750-752.
5. Simon H.A. Problem Formulation and Alternative Generation in the Decision Making Process // Progress in Decision, Utility and Risk Theory / Ed. Chikan A., et al. Boston, MA: Kluwer, 1991. P. 77-84.
6. Simon H. Reason in Human Affairs. Stanford University Press, 1983. P. 115.
7. Lemieux M., Bordage G. Professional Versus Structural Semantic Analysis of Medical Diagnostic Thinking // Cognitive Science. 1992. No. 16. P. 198-204.
8. Larichev O.I., Bolotov A.A. The DIFKLASS System: Construction of Complete and Noncontradictory Expert Knowledge Bases in Problems of Differential Classification // Automatic Documentation and Mathematical Linguistics. 1996. V.30, No. 5. P. 12–22.
9. Болотов А.А., Барчуков В.Г., Тен А.М., Онопченко О.В. Оценка информативности признаков в системе интеллектуальной поддержки тренера в циклических видах спорта // День спортивной информатики: Материалы III-й научно-практической конференции. М., 2019. С. 140-143.
10. Болотов А.А., Ларичев О.А. Сравнение методов распознавания образов по точности аппроксимации разделяющих гиперплоскостей // Автоматика и телемеханика. 1995. № 7. С. 116-123.

REFERENCES

1. Larichev O.I., Mechitov A.I., Moshkovich Ye.M., Furems Ye.M. Vyyavleniye Ekspertnykh Znaniy = Identification of Expert Knowledge. Moscow, Nauka Publ., 1989. 128 p. (In Russ.).
2. Larichev O. I. Theory of Subconscious Decision Rules and Its Application in Diagnostic Tasks. Psikhologicheskiy Zhurnal = Psychological Journal. 2003;24;1:56-64 (In Russ.).
3. Kilstrom J. The Cognitive Unconscious. Science. 1987;237;4821:1445-1452.
4. Larichev O.I. Structures of Expert Knowledge in Classification Problems. Doklady Akademii Nauk. 1994;336;6:750-752 (In Russ.).
5. Simon H.A. Problem Formulation and Alternative Generation in the Decision Making Process. Progress in Decision, Utility and Risk Theory. Ed. Chikan A., et al. Boston, MA, Kluwer, 1991. P. 77-84.
6. Simon H. Reason in Human Affairs. Stanford University Press, 1983. P. 115.
7. Lemieux M., Bordage G. Professional Versus Structural Semantic Analysis of Medical Diagnostic Thinking. Cognitive Science. 1992;16:198-204.
8. Larichev O.I., Bolotov A.A. The DIFKLASS System: Construction of Complete and Noncontradictory Expert Knowledge Bases in Problems of Differential Classification. Automatic Documentation and Mathematical Linguistics. 1996;30;5:12–22.
9. Bolotov A.A., Barchukov V.G., Ten A.M., Onopchenko O.V. Evaluation of Informative Features in the System of Intellectual Support of a Coach in Cyclic Sports. Den Sportivnoy Informatiki. = Sports Informatics Day. Materials of the III Scientific and Practical Conference. Moscow Publ., 2019. P. 140-143 (In Russ.).
10. Bolotov A.A., Larichev O.I. Comparison of Pattern Recognition Methods by the Accuracy of Approximation of Separating Hyperplanes. Avtomatika i Telemekhanika = Automation and Remote Control. 1995;7:116-123 (In Russ.).

Конфликт интересов. Авторы заявляют об отсутствии конфликта интересов.
Финансирование. Исследование не имело спонсорской поддержки.
Участие авторов. Статья подготовлена с равным участием авторов.
Поступила: 20.09.2022. **Принята к публикации:** 20.10.2022.

Conflict of interest. The authors declare no conflict of interest.
Financing. The study had no sponsorship.
Contribution. Article was prepared with equal participation of the authors.
Article received: 20.09.2022. **Accepted for publication:** 20.10.2022