

УДК 004.891.3

К.В. Мельник, С.И. Ершова

Национальный технический университет «ХПИ», Харьков

ПРОБЛЕМЫ И ОСНОВНЫЕ ПОДХОДЫ К РЕШЕНИЮ ЗАДАЧИ МЕДИЦИНСКОЙ ДИАГНОСТИКИ

Рассмотрены основные проблемы, возникающие при решении задачи медицинской диагностики. Предлагаются подходы формализации медицинских данных для задачи диагностирования. Приведена классификация методов обработки данных и их основные особенности при решении задачи медицинской диагностики.

Ключевые слова: медицинская диагностика, представление знаний, семантические сети, фреймы, онтологии, нейросетевые технологии.

Введение и актуальность задачи

Развитие системы здравоохранения и выход Украины на европейский уровень в медицине происходит достаточно медленно, чему способствует целый ряд проблем.

Основная проблема – это недостаток финансирования, что неизбежно отражается на уровне медицинской техники и лабораторного оборудования, используемых технологий и методов диагностики. Кроме этого, низкий уровень оплаты труда приводит к уходу квалифицированных специалистов в другие сферы деятельности. С другой стороны, ограниченное финансирование влияет на качество подготовки молодых врачей, развитие медицинской науки и последипломного образования.

Актуальной проблемой здравоохранения является тенденция к увеличению риска различных заболеваний, к повышению уровня смертности, что связано с ухудшающейся экологией, низким уровнем жизни, несвоевременной диагностикой, а также недостаточной профилактикой различных нарушений здоровья.

Использование в медицинских учреждениях современных информационных компьютерных технологий позволяет не только повысить эффективность процессов обработки медицинской информации, но и открывает большие перспективы в решении многих проблем современной медицины. К настоящему моменту создание и использование медицинских СППР (систем поддержки принятия решений) выделилось в отдельное направление научных исследований, которое объединяет результаты, полученные для интеллектуальных систем, в частности ЭС (экспертных систем) [1, 14, 16], подходы Data Mining [1], а также современные концепции информационных технологий [1].

Одна из основных задач, решению которой посвящено большое число публикаций, связана с созданием систем для медицинской диагностики.

Обработке медицинской информации и последующему принятию решений посвящены, например, такие работы: [1, 5 – 7, 13, 15]. Несмотря на это, еще достаточно много вопросов в данной области осталось открытыми, что отмечают, в частности, авторы [3, 4, 9].

Цель данного исследования – провести анализ существующих подходов к решению задач медицинской диагностики и оценить их с точки зрения реализации медицинских СППР.

Представление знаний в интеллектуальных системах

Сфера медицинского диагностирования является достаточно сложной предметной областью ввиду ее особенностей описания, не всегда интуитивно понятных связей между признаками и вообще учета того, что человеческий организм всегда непредсказуем. На сегодняшний день можно выделить такие модели представления знаний, как семантические сети, фреймы, логические модели, онтологии [1, 10, 14].

Семантические сети – модель представления знаний, которая наиболее близка к естественному языку [14]. Семантическая сеть представляет собой ориентированный граф с вершинами, которым соответствуют объекты, понятия или ситуации, и дугами, которые могут быть определены различными методами, характеризующими отношения между объектами. К достоинствам семантических сетей можно отнести большие выразительные возможности; наглядность системы знаний, представленной графически; близость к естественному языку; соответствие современным представлениям об организации долговременной памяти человека; легкое настраивание. Отрицательными моментами использования сетевой модели являются такие факты: эта модель не дает ясного представления о структуре предметной области, которая ей соответствует, поэтому формирование и модификация такой модели затруднительны; сетевые модели представляют со-

бой пассивные структуры, для обработки которых необходим специальный аппарат формального вывода и планирования; сложность поиска и вывода на семантических сетях; наличие множественных отношений между элементами сети.

По своей организации фреймовая модель во многом напоминает семантическую сеть. Фрейм – это совокупность узлов и отношений, организованных иерархически, где верхние узлы представляют собой общие понятия, а нижние узлы – более частные случаи этих понятий [14]. Фреймы отображают типичную схему действий в реальной ситуации. Понятие фреймов первым ввел Минский [20]. Каждый узел представляет собой атрибут – слот, содержащий какое-то определенное понятие и значение этого понятия, либо в качестве значения слота может выступать имя другого фрейма, либо слот может содержать имя процедуры, с помощью которой можно вычислить определенное значение, либо в слоте может находиться диапазон значений. Совокупность фреймов можно объединить в сеть, при этом свойства фреймов наследуются сверху вниз, самый верхний фрейм содержит более общую информацию, которая является справедливой для всей иерархии фреймов. Использование фреймов при формализации предметной области дает достаточно эффективные результаты [14]. К отрицательным моментам при использовании фреймовой модели представления знаний можно отнести трудности адаптации модели при поступлении новых знаний, достаточно большие объемы памяти, необходимые для хранения элементов модели, также отсутствие механизмов управления выводом.

Основная идея построения логических моделей представления знаний состоит в том, что вся информация, необходимая для решения прикладных задач, рассматривается как совокупность фактов и утверждений, которые представляются как формулы в некоторой логике [1, 4]. Знания отображаются совокупностью таких формул, а получение новых знаний сводится к реализации процедур логического вывода. К положительным характеристикам логических моделей можно отнести тот факт, что в качестве основы используется классический аппарат математической логики, методы которой достаточно хорошо изучены и формально обоснованы [14]; существуют достаточно эффективные процедуры вывода [3, 6]; в базах знаний можно хранить лишь множество аксиом, а все остальные знания получать из них по правилам вывода. Основными недостатками логического метода представления знаний являются значительные затраты времени на построение цепочки вывода; невозможность эффективно описать правила с исключениями; необходимость описания большого числа правил и утверждений при моделировании реальной системы, которые могут быть противоречивыми.

Онтологии – это еще один способ описания предметной области, базовых понятий этой области, их свойств и связей между ними. Практически все модели онтологий содержат концепты (понятия, классы, сущности, категории), свойства концептов (слоты, атрибуты, роли), отношения между концептами (связи, зависимости, функции) и дополнительные ограничения [10]. Онтология вместе с множеством отдельных экземпляров составляет базу знаний, описывающую факты на основе общепринятого смысла используемого словаря. Преимущества онтологий определяются возможностью совместного использования людьми или программными агентами для общего понимания структуры информации; возможностью повторного использования знаний в предметной области; возможностью отделения знаний в предметной области от оперативных данных; возможностью анализа знаний в предметной области.

В качестве примера использования модели семантической сети для решения задач медицинской диагностики можно привести систему CASNET (Causal-Associational NETWORK) [16]. На основе логических моделей построены такие системы как MYCIN – система для диагностирования бактериальных инфекций или система INTERNIST – система диагностирования в области общей терапии [1, 16]. Учитывая уровень развития информационных компьютерных технологий, онтологии нашли свое применение во многих предметных областях, в частности, в медицине можно отметить словарь SNOMED [10] или Систему унифицированного медицинского языка (the Unified Medical Language System), которые используются для формализации общей терминологии и аннотирования в области медицины.

Методы обработки медицинской информации

Формально задачу медицинского диагностирования можно представить как задачу классификации, которая состоит в том, что чтобы поставить в соответствие совокупности входных параметров конкретное заболевание [3, 13].

Основные подходы, которые применяются для решения задачи медицинской диагностики можно сгруппировать следующим образом:

1. Логический подход.
2. Статистический подход.
3. Бионический подход.

Логический подход при принятии решений в медицине является довольно распространенным, так как является прямым отображением рассуждений врача [3]. Рассуждения врача во время диагностического процесса должны быть определенными, последовательными и доказательными. То есть для определения того или иного нарушения, которое характеризуется комплексом симптомов необходимо исполь-

зовать законы формальной логики. Можно поставить диагноз, используя индуктивные рассуждения, связанные с прогнозированием результатов наблюдений на основе прошлых опытов. Но индуктивный метод в диагностике несостоятелен, так как логика познания и рассуждений носит дедуктивный характер. Поэтому обычно используется дедуктивный метод при определении диагноза, то есть процесс дедуктивного вывода из некоторого множества подозреваемых заболеваний необходимого диагноза.

Разновидностью логического вывода является метод деревьев решений или деревьев классификаций [4]. Для использования данного подхода составляется база правил вида: "Если ...То" в виде иерархической структуры, которая представляет собой дерево. Для определения класса заболевания необходимо ответить на вопросы, стоящие в узлах этого дерева, начиная с его корня, и таким образом осуществить переход к последующему вопросу. Положительными сторонами данного подхода можно считать наглядность метода и понятность. Но на практике возникает ряд вопросов, которые ограничивают использование данного подхода при решении задачи классификации. Одна из проблем связана с выбором следующего узла. Для ее решения используются различные алгоритмы, которые зачастую дают слишком детализированную структуру дерева и могут приводить к ошибкам. Несмотря на различные особенности данного подхода, логическая схема вопросов-ответов диагностического процесса в виде деревьев решений успешно применяется на практике [3, 4].

К группе статистических методов можно отнести байесовский подход, методы дискриминантного анализа, вывод, основанный на прецедентах. Примеры использования данных методов для решения задачи медицинского диагностирования приведены, например, в работах [1, 3, 4]. Использование теоремы Байеса при определении класса заболевания – довольно распространенный подход ввиду своей простоты, наглядности и простых математических вычислений, но у него есть явный недостаток – необходима большая база архивных данных для того, чтобы вероятность того или иного диагноза соответствовала реальности. Также отрицательным моментом является тот факт, что редкое заболевание либо нестандартные симптомы трудно правильно идентифицировать [3].

Дискриминантный анализ характеризуется наличием большого количества вычислений, выявление различного рода связей между признаками, выяснением влияния связей признаков на результат диагностирования, что обычно влечет к трудностям при решении поставленной медицинской задачи [1].

Метод вывода, основанный на прецедентах (Case Based Reasoning) – это подход, основанный на

использовании предыдущих случаев и опыта [4, 19]. Для принятия решения вначале осуществляется поиск аналогичных прецедентов, которые имели место в прошлом, а затем вычисляется мера близости между новым и всеми найденными случаями, для которых определены варианты решений. У данного подхода есть ряд ограничений: система, основанная на прецедентах, построена на знаниях, которые "извлекаются" из экспертов, значит, присутствует мера субъективизма; знания, полученные от людей, необходимо оценить, проверить и определить их достоверность; отсутствие решения таких задач, для которых нет прецедентов; произвол в выборе меры близости; резкое снижение производительности при большом наборе входных параметров. Примеры реализации данного подхода к решению задачи медицинского диагностирования рассматриваются, например, в работах [4, 17, 18].

В работе [4] приведен еще один пример реализации статистического подхода – метод фазового интервала, который основан на идее, что состояние организма и его функции можно описать при помощи системы бинарных параметров, где каждый параметр может принимать два значения: ноль – отсутствие признака или единицу – наличие признака. Тогда любое состояние пациента можно представить в виде точки в фазовом пространстве. Заболевания одной группы объединяются в область, для которой вычисляются координаты центра множества – точка области, соответствующая наиболее типичному состоянию для данной группы заболеваний, а также диаметр области – максимальное расстояние между двумя точками. Диагноз устанавливается по минимальному расстоянию между центрами областей и рассматриваемым заболеванием. Расстояние между двумя точками представляет собой число несовпадений признаков. Для увеличения эффективности принятия решений при диагностическом процессе и минимизации субъективности априорной информации каждому признаку можно придать вес, который впоследствии необходимо использовать при вычислении расстояния. Трудность данного процесса состоит в том, что для каждого заболевания свои информативные признаки, а поэтому веса необходимо переопределять для каждой группы заболевания.

Бионический подход представляет собой процесс искусственного воспроизведения тех структур и процессов, которые характерны для человеческого мозга. В рамках данного подхода для решения задачи диагностирования было разработано нейросетевое направление [2, 12]. Достоинства, присущие данному подходу весьма обширны: способность к адаптации (обучению и самообучению); параллельность обработки информации; робастность (устойчивость к отдельным сбоям). Нейронные сети (НС)

используют при решении задач медицинской диагностики, задач классификации (кластеризации), аппроксимации, прогнозирования [2, 12]. Иногда применимость нейросетевого подхода ограничена ввиду некоторых недостатков: необходимо большое количество архивных данных для обучения созданной нейронной сети; большой вес имеет фактор субъективности при создании структуры НС – количество слоев, количество нейронов в каждом слое и функция активации выбираются экспертом, что, в свою очередь, вносит дополнительную неопределенность; также при обучении НС существует вероятность того, что выбранная структура не даст адекватные результаты при использовании некоторых входных данных; достаточная чувствительность к входным данным – тренировочный шаблон, на котором обучается НС, должен быть соответствующим образом подготовлен для элиминации дополнительного веса отдельных входных параметров, приводящего к недостоверным результатам обучения. В работах [2, 9, 11, 12] представлены примеры применения нейросетевого подхода при решении различного рода задач медицинского диагностирования.

Таким образом, задача медицинской диагностики представляет собой достаточно сложную задачу ввиду того, что данные о пациенте являются слабо структурированными и имеют различный характер. Часть необходимой информации, касающейся пациента, обычно отсутствует, что вносит дополнительные трудности при обработке медицинских данных; часть информации носит качественный характер, так как ее определяет врач, то есть присутствует доля субъективизма; часть информации отражает результаты анализов, а значит, необходимо учитывать фактор случайности из-за ошибок измерений. На сегодняшний день сложилось несколько подходов работы с неопределенностью в задачах медицинского диагностирования. Вероятностный подход представляет собой подход, когда неизвестные факторы статистически устойчивы и поэтому представляют собой случайные величины или случайные события [4]. При этом должны быть определены все необходимые статистические характеристики: законы распределения и их параметры, функции или плотности распределения вероятностей, что, в свою очередь, вносит дополнительные трудности.

Другим подходом учета неопределенности является использование теории нечетких множеств, которая позволяет учесть недоступные или неизвестные данные, или данные, носящие исключительно качественный характер [6, 8, 12]. Однако применение данного подхода связано с рядом трудностей, например, с определением вида функции принадлежности, а также трудностями последующей обработкой нечетких данных.

Вывод

Задача медицинской диагностики – сложная и многогранная задача, решением которой занималось не одно поколение ученых. На сегодняшний день разработано большое количество подходов к ее решению, но каждый из них имеет свои недостатки и ограничения, поэтому необходимость разработки эффективных моделей и методов медицинской диагностики не утратила своей актуальности.

Список литературы

- 1 Дюк В. *Обработка данных на ПК в примерах* / В. Дюк. – СПб: Питер, 1997. – 240 с.
- 2 Круглов В.В. *Искусственные нейронные сети. Теория и практика* / В.В. Круглов, В.В. Борисов. – 2-е изд., стереотип. – М.: Горячая линия – Телеком, 2002. – 382 с.
- 3 Ластед Л. *Введение в проблему принятия решений в медицине* / Л. Ластед; пер. с англ. И.М. Быховского. – М.: Мир, 1971. – 283 с.
- 4 Мисюк Н.С. *Диагностические алгоритмы* / Н.С. Мисюк, А.М. Гурленя, В.В. Лозовик. – Мн.: Вышэйшая школа, 1970. – 187 с.
- 5 Мацуга О.М. *Інформаційна технологія обробки неоднорідних медичних даних для підтримки прийняття рішень під час діагностики: дис. ... канд. техн. наук: 05.13.06* / Мацуга Ольга Миколаївна. – Д.: Дніпропетровський національний ун-т, 2007. – 209 с.
- 6 Мельник К.В. *Процедура диагностирования состояния сердечно-сосудистой системы пациента на основе нечеткой логики* / К.В. Мельник, А.Е. Голоскоков // *Вестник НТУ «ХПИ»*. – X, 2008. – № 49. – С. 101-104.
- 7 Мельник К.В. *Особенности обработки данных для медицинской экспертной системы* / К.В. Мельник // *Інформаційні технології: наука, техніка, технологія, освіта, здоров'я: тези доповідей міжнародної науково-практичної конференції*. – X.: НТУ «ХПИ», 2010. – Ч. 1. – С. 15.
- 8 Мельник К.В. *Разработка системы управления лечением пациента, основанной на нечеткой ситуационной сети* / К.В. Мельник // *Інформаційні технології та інформаційна безпека в науці, техніці та освіті «Інфотех – 2007»: матеріали міжнародної науково-практичної конференції*. – Севастополь, 2007. – Ч. 2. – С. 13-17.
- 9 *Распознавание образов и медицинская диагностика* / Ю.И. Неймарк, З.С. Баталова, Ю.Г. Васин, М.Д. Брейдо. – М.: Наука, 1972. – 328 с.
- 10 *Онтологии и тезаурусы: учеб. пособие* / В.Д. Соловьев, Б.В. Добров, В.В. Иванов, Н.В. Лукашевич. – Казань – Москва, 2006. – 173 с.
- 11 Ежов А. *Нейронные сети в медицине* / А. Ежов, В. Четкин // *Открытые системы*. – К., 1997. – № 4. – С. 34-37.
- 12 Рутковская Д. *Нейронные сети, генетические алгоритмы и нечеткие системы* / Д. Рутковская, М. Пилиньский, Л. Рутковский; пер. с польского И.Д. Рудинского. – 2-е изд. – М.: Горячая линия-Телеком, 2008. – 452 с.
- 13 Реброва О.Ю. *Статистический анализ медицинских данных. Применение пакета прикладных программ STATISTICA* / О.Ю. Реброва. – М.: Медиасфера, 2002. – 312 с.
- 14 Уотермен Д. *Руководство по экспертным системам: пер. с англ.* / Д. Уотермен. – М.: Мир, 1989. – 388 с.
- 15 Хачатурідзе Т.М. *Інформаційна технологія підтримки прийняття рішень в системі моніторингу при діагностиці захворювань: дис. ... канд. техн. наук:*

05.13.06 / Хачатурідзе Тетяна Миколаївна. – Д.: Дніпропетровський національний ун-т, 2006. – 173 с.

16 Элти Дж. Экспертные системы: концепции и примеры: пер. с англ. / Дж. Элти, М. Кумбс. – М.: Финансы и статистика, 1987. – 191 с.

17 Abdel-Badeeh M. Salem. Case Based Reasoning Technology for Medical Diagnosis / Abdel-Badeeh M. Salem // World Academy of Science, Engineering and Technology. – 2007. – № 31. – P. 9-13.

18 Abdel-Badeeh M. Salem. A Case Based Expert System for Supporting Diagnosis of Heart Diseases / Abdel-Badeeh M. Salem, Mohamed Roushdy, Rania A. HodHod. //

ICGST International Journal on Artificial Intelligence and Machine Learning. – 2005. – P. 33-39.

19 Janet L. Kolodner. An Introduction to Case-Based Reasoning / Janet L. Kolodner // Artificial Intelligence Review. – 1992. – № 6. – P.3-34.

20 Minsky M. A Framework for Representing Knowledge / M. Minsky // MIT-AI Laboratory Memo 306, 1974.

Поступила в редколлегию 17.02.2011

Рецензент: д-р техн. наук, проф. М.Д. Годлевский, Национальный технический университет «ХПИ», Харьков.

ПРОБЛЕМИ ТА ОСНОВНІ ПІДХОДИ ДО РІШЕННЯ ЗАДАЧІ МЕДИЧНОЇ ДІАГНОСТИКИ

К.В. Мельник, С.І. Єршова

Розглянуто основні проблеми, що виникають при розв'язанні задачі медичної діагностики. Пропонуються підходи формалізації медичних даних для завдання діагностування. Наведено класифікацію методів обробки даних і їх основні особливості при розв'язанні задачі медичної діагностики.

Ключові слова: медична діагностика, представлення знань, семантичні мережі, фрейми, онтології, нейромережеві технології.

PROBLEMS AND BASIC APPROACHES TO SOLVING THE PROBLEM OF MEDICAL DIAGNOSIS

K.V. Melnik, S.I. Ershova

Basic problems, arising up at the decision of task of medical diagnostics, are considered. Approaches of formalization of medical information are offered for the task of diagnosing. Classification of methods of processing of data and their basic features is resulted at the decision of task of medical diagnostics.

Keywords: medical diagnostics, representation of knowledges, semantic networks, frames, to ontology, connectionist technology.