

МИНОБРНАУКИ РОССИИ

Федеральное государственное бюджетное образовательное учреждение
высшего образования

**«САРАТОВСКИЙ НАЦИОНАЛЬНЫЙ ИССЛЕДОВАТЕЛЬСКИЙ
ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ
ИМЕНИ Н. Г. ЧЕРНЫШЕВСКОГО»**

Кафедра дискретной математики и информационных технологий

**МЕТОДЫ РАСПОЗНАВАНИЯ ОБРАЗОВ В ДИАГНОСТИКЕ
РЕВМАТИЗМА У ВЗРОСЛЫХ И ДЕТЕЙ**

АВТОРЕФЕРАТ БАКАЛАВРСКОЙ РАБОТЫ

Студентки 4 курса 421 группы
направления 09.03.01 — Информатика и вычислительная техника
факультета КНиИТ
Кабделовой Эльвиры Тимиргалиевны

Научный руководитель
доцент, к.ф.-м.н.

Ю.А. Бродская

Заведующий кафедрой
доцент, к. ф.-м. н.

Л. Б. Тяпаев

Саратов 2021

ВВЕДЕНИЕ

В рамках кибернетики во второй половине 50-х годов XX в. начало формироваться новое научное направление, связанное с разработкой теоретических основ и практической реализацией устройств, а затем и систем, предназначенных для распознавания неизвестных объектов, явлений, процессов. Новая научная дисциплина получила название «Распознавание образов». Подобное название возникло в связи с тем, что процесс распознавания отождествляется с выявлением вопроса о том, к какому классу объектов (образу) может быть отнесен распознаваемый объект. При этом класс олицетворяет собой некоторую совокупность (подмножество) объектов, обладающих близкими свойствами.

В качестве предметной области выбрана медицина, а именно ее раздел - ревматология. Ревматизм – это системное заболевание соединительной ткани преимущественным поражением сердечно-сосудистой системы, развивающееся в связи с острой стрептококковой инфекцией у предрасположенных к нему лиц в возрасте 7-15 лет. В настоящее время количество больных ревматизмом с поражением суставов уменьшилось, а число больных в зрелом возрасте с поражением сердца возрастает, так как болезнь проходит бесприступно, и в подростковом возрасте часто ошибаются при дифференциальной диагностике. Ревматизм имеет схожий симптомокомплекс со многими другими заболеваниями (когда клинические проявления очень близки). В связи с этим специалисты испытывают сложности в прогнозировании данной нозологической формы. О наличии этого заболевания узнают в зрелом возрасте, когда вероятность выздоровления очень мала. Актуальность данной работы заключается в необходимости помощи медицинским работникам для уменьшения вероятности ошибок при диагностике и минимизации риска летального исхода. С помощью математических методов распознавания образов необходимо реализовать программу, благодаря которой можно уточнить диагноз, и тем самым облегчить процесс диагностирования специалисту. В данной работе рассматриваются и выбираются для реализации математические методы распознавания образов для диагностики ревматизма у взрослых и детей. Вопросами дифференциальной диагностики ревматизма занимались: Неймарк Ю.И., Баталов З.С и Рогов Д.А. Основной целью работы является выбор и реализация математического метода диагностики ревматизма в среде про-

граммирования Visual Studio 2019 на языке C++. Для достижения цели поставлены следующие задачи:

1) изучение литературных источников, рассматривающих математические модели распознавания образов и описывающих выбранную предметную область (т.е. литературных источников по диагностике ревматизма);

2) формирование признаков и получение решающего правила для диагностирования с помощью обучающего статистического материала;

3) изучение математических методов распознавания образов: минимаксного критерия и тестового;

4) сравнение минимаксного и тестового метода, выбор одного для реализации;

5) реализация метода минимаксного критерия на языке C++ для диагностики ревматизма.

1 Краткое содержание работы

Выпускная квалификационная работа состоит из введения, 7 глав, заключения, 11 таблиц, 7 рисунков, 1 приложения и 22 источников литературы. В введении описывается актуальность темы, цели и задачи выпускной работы.

В 1 главе дано описание основных понятий теории распознавания образов. Распознавание образов (объектов) – это задача идентификации объекта по его изображению (оптическое распознавание), аудиозаписи (акустическое распознавание) или другим характеристикам. Образ – это классификационная группировка, которая позволяет объединить группу объектов по некоторым признакам.

Одним из базовых определений также является и понятие множества. В компьютере множество - это набор неповторяющихся однотипных элементов. «Неповторяющихся» - значит, что элемент в множестве либо есть, либо нет. Универсальное множество включает все возможные элементы, пустое не содержит ни одного.

Методика отнесения элемента к какому-то образу называется решающим правилом. Еще одно важное понятие - метрика - определяет расстояние между элементами множества. Чем меньше это расстояние, тем больше схожи объекты (символы, звуки и др.), которые мы распознаем.

Во 2 главе описан обзор и дан выбор математического метода распознавания образа. Изучение таких методов как: геометрического, вероятностного, тестового, минимаксного критерия.

Геометрический метод распознавания основан на использовании некоторой функции подобия (принадлежности) S объекта данному классу. Эта функция определяет некоторую меру близости объекта b_j с координатами $x = (x_1, x_2, \dots, x_N)$ к множеству эталонов

$$y^m = (y_1^m, y_2^m, \dots, y_N^m).$$

Решающее правило для геометрического распознавания будет формулироваться следующим образом. Пусть даны два класса A_1 и A_2 . Для распознавания принадлежности какого-нибудь объекта p к классу x_m определяется

величина

$$\begin{aligned}
 S(p, \{x_m\}) &= \frac{1}{M} \sum_{m=1}^M d^2(p', x'_m) = \frac{1}{M} \sum_{m=1}^M \sum_{n=1}^N (p'_n - p'_{mn})^2 = \\
 &= \frac{1}{M} \sum_{m=1}^M \sum_{n=1}^N \left[\sum_{s=1}^N a_{ns} (p_s - x_{ms}) \right]^2.
 \end{aligned}$$

Решающее правило состоит в следующем:

$$\begin{cases} p \in A_1, & \text{если } S_{A_1}(p, \{x_m\}) < S_{A_2}(p, \{x_m\}); \\ p \in A_2, & \text{если } S_{A_1}(p, \{x_m\}) > S_{A_2}(p, \{x_m\}), \end{cases}$$

причем в каждой из функции S_{A_1} и S_{A_2} коэффициенты a_{ns} ищутся, применительно к первому или второму множеству эталонов.

Тестовый метод

В данном алгоритме решающее правило выглядит следующим образом. Для каждой строки $s = (s_1, \dots, s_n)$ матриц T_1 и T_2 находим скалярное произведение

$$s * p = \sum_{i=1}^n s_i * p_i$$

называемое весом строки. Пусть q — вес объекта $x = (x_1, \dots, x_n)$, подлежащего классификации. Тогда, если $q \geq h$, то отнесем x к классу K_1 а иначе — к классу K_2 . Значение h подбирается с учетом требования, чтобы наибольшее число строк матриц T_1 и T_2 классифицировались правильно. Разделяющей поверхностью в алгоритме A_2 является гиперплоскость, заданная уравнением

$$\sum_{i=1}^n x_i * p_i = h$$

Идея тестового подхода основывается эвристике голосования теста. Если дан объект x , который надо распознать, и некоторый тест τ , то отбрасываются все признаки, которые не входят в тест. Если x на множестве оставшихся признаков больше похож на элементы обучающей выборки из первого класса, то говорят, что тест τ голосует за первый класс, иначе — за второй. Далее выбирается некоторое опорное множество тестов, в качестве которого можно

взять множество всех тестов, множество тупиковых тестов, или множество тестов фиксированной длины, и вычисляется, за что интегрально голосует опорное множество.

Вероятностный метод

Этот метод основан на теореме Байеса из теории вероятностей, которая применительно к задаче распознавания может быть сформулирована следующим образом. Имеется полная группа несовместных гипотез, роль которых при распознавании выполняют образы:

$$A : A_1, A_2, \dots, A_i, \dots, A_M.$$

Байесовский критерий распознавания основан на сравнении апостериорных вероятностей принадлежности объекта b_j классу A_i . После предъявления объекта эти вероятности подсчитываются по формуле (??). Сравнивая все значения $P(A_i/b_j)$, относим объекта к такому классу, для которого это значение максимально. Рассмотрим для простоты случай наличия двух классов, $M = 2$. Для распознавания необходимо иметь в машине условные вероятности распределения события, состоящего в появлении объекта b_j со значениями признаков $x = (x_1, x_2, \dots, x_k, \dots, x_N)$ при условии, что объект относится к классу A_i :

$$P(x_1, x_2, \dots, x_k, \dots, x_N/A_1) = P_{A_1}(x).$$

Известна также плотность распределения вероятности $P_{A_2}(x)$ события, состоящего в том, что предъявляемый объект относится к классу A_2 .

Минимаксный критерий

При построении систем классификации (расознавания) возможна ситуация, когда априорные вероятности появления объектов соответствующих классов неизвестны. Поэтому минимизировать значение среднего риска принятия решения на основании байесовской стратегии не представляется возможным. В этой ситуации используется критерий, который минимизирует максимально возможное значение среднего риска. Этот критерий называется минимаксным критерием [2]. Среди всевозможных решающих правил распознавания и классификации линейные решающие правила выгодно выделяются своей простотой. С помощью данного правила решение задачи ме-

дицинской диагностики допускает широкое практическое использование.

В качестве требования к правилу классификации выдвигается условие его минимаксности, состоящее в требовании минимальности максимальной вероятности ошибки, которая может произойти из-за погрешностей описания [5].

В 3 главе рассказано о кодировании исходных данных. Вопрос о выборе признаков для постановки диагноза и их кодирования непрост. При кодировании признаков область изменения каждого признака разбивается на несколько градаций, например, при дифференциальной диагностике ревматизма были приняты такие градации: 1) менее 20 лет, 2) 20-29 лет, 3) 30-39 лет, 4) 40 лет и более.

В 4 главе описывается формальная постановка задачи.

В 5 главе дано описание процесса решения задач в медицинской диагностике. Математические методы распознавания образов позволяют найти решающее правило получения π по x и даже часто, когда врачебная диагностика затруднительна.

В 6 главе описывается исследование бесприступно текущего ревматизма у детей. В таблицах приведены результаты дифференциальной диагностики.

В 7 главе представлена программная реализация минимаксного критерия. В среде программирования Visual Studio 2019 на языке C++ реализована программа с помощью минимаксного критерия для минимальности максимальной вероятности ошибки. И приведен тестовый пример пациента.

ЗАКЛЮЧЕНИЕ

В ходе написания работы реализован математический метод (минимаксный критерий) в среде программирования Visual Studio 2019 на языке C++. Для осуществления поставленной заявленной в работе цели была проделана следующая работа:

- 1) изучены литературные источники Журавлева Ю.И., Кудрявцева В.Б., Ф.Розенבלата, Богомоллова А.М., а также Вапника В.Н. и С. В. Яблонского по математическим моделям распознавания образов, а также предметной области (медицинская литература по диагностике ревматизма);
- 2) опираясь на данные, полученные в медицинских исследованиях, сформированы признаки для диагностики;
- 3) изучены и детализированы такие методы как: минимаксный критерий и тестовый подход;
- 4) подобраны тестовые примеры;
- 5) сделан выбор среди двух методов в пользу минимаксного критерия;
- 6) сделана программная реализация на языке C++;
- 7) составлен печатный отчет по проделанной работе.

В результате проделанной работы можно сделать следующие выводы. Вероятностный метод осуществлялся с помощью минимаксного критерия. Использование такого метода дает улучшение качества диагностики. Из 706 больных[3] общее число несовпадений с верифицированным диагнозом врача возросло до 12 случаев, а число отказов увеличилось до 38. В классе больных ревматизмом ошибка составила 2,5% при 7,8% отказов. В классе прочих заболеваний соответственно получено 4,0% и 8,0% отказов. Благодаря методам математического распознавания образов и их программной реализации, можно получить результаты дифференциальной диагностики ревматизма близкими к диагностированию заболеванию специалистом.

СПИСОК ИСПОЛЬЗОВАННЫХ ИСТОЧНИКОВ

- 1 Горелик А.Л. Скрипкин В.А. Методы распознавания. Учеб. пособие. - 2-е издан. Выск. шк.. 1984 - 208 с.ил.
- 2 Степанова М. Д. , Самодумкин С. А., Лемешева Т. Л., Математические методы диагностики в медицинских интеллектуальных системах, Мн.: БГУИР, Минск, 2001.
- 3 Распознавание образов и медицинская диагностика / Под ред. Ю. Неймарка. М.: Наука, 1972.
- 4 Кудрявцев В.Б., Интеллектуальные системы Издательство Юрайт, 2016
- 5 Гасанов Э. Э. Некоторые оценки сложности поиска информации // Физическое и математическое моделирование дискретных систем. Межвузовский сборник трудов № 56. — М.: Изд-во Моек, энерг. ин-та, 1985. — С. 43-47.
- 6 Балабан В. Г., Мощич П.С., Некоторые вопросы дифференциальной диагностики тонзиллогенной интоксикации и вяло текущего ревматизма у детей. Вопросы ревматизма у детей, №4, 1963
- 7 Буркина З.П., Медихова Н.И. К вопросу о диагностике ревматизма у детей. Педиатрия, №12, 1964.
- 8 Воловик А.Б., Об избыточной диагностике ревматизма у детей. В сб. «Ревматизм у детей», Ленинград, 1959, стр. 381-387.
- 9 Талызина Н. К. К клинической характеристике латентной формы ревматизма у детей. Сб. «Вопросы клиники и профилактики ревматизма у детей», Омск, 1967, стр.5-10.
- 10 Гатаулина Л.Д. Распространение ревматизма в организованных детских коллективах. Вопросы охраны материнства и детства, №12, 1965.
- 11 Себестиан Г.С., Процессы понятия решений при распознавании образов, Изд-во «Техника», Киев, 1965.
- 12 Сафронов В. В. Распространение ревматизма среди детей дошкольного возраста в г. Горьковском и Горьковской области. Педиатрия, № 2, 1966.
- 13 Якубович В. А. Машины, обучающиеся распознаванию образов. Сб. «Методы вычислений», вып. 2, Изд-во ЛГУ, 1963.

- 14 Неймарк Ю. И., Баталова З., Образцова Н.Д., К вопросу о выборе признаков при распознавании образов, Изв. АН СССР, Техн. Кибернетика, № 1, 1970, стр. 105-112.
- 15 Любавин Я.К., Кочкин В.И., Неймарк Ю.И., Баталова З.С., Образцова Н.Д. Дифференциальный диагноз ревматизма и тонзилло-кардиального синдрома у детей с помощью ЭВМ. Сб. «Вопросы охраны материнства и детства», №3, Изд-во «Медицина», 1968.
- 16 Насонова В.А, Кузьмина Н.Н. Ревматизм. – В кн.: «Руководство по ревматическим болезням», М.: Медицина, 1997, с. 144 – 160.
- 17 Журавлев Ю. И., Гуревич И. Б. Распознавание образов и распознавание изображений / Распознавание, классификация и прогноз. М.: Наука, 1989, вып. 2.
- 18 Соловьев Н.А. Тесты: (теория, построение, применение). Новосибирск: Наука, 1978.- 189 с.
- 19 Яблонский С.В. О тестах для электрических схем. – «Усп. Матем. наук», 1955, т.10, вып.4 (66), с. 182-184
- 20 Вапник В. Н., Червоненкис А.Я., Теория распознавания образов, Изд-во «Наука», 1974.
- 21 Кудрявцев В. Б., Андреев А.Е., Гасанов Э.Э. Теория тестового распознавания // М: изд-во «Физматлит», 2007, 320 с.
- 22 Чегис И.А., Яблонский С.В. Логические способы контроля электрических схем.// Труды математического института им. В.А. Стеклова. 1958. Т.51.- С.226-269.