

Министерство образования и науки Российской Федерации

ФЕДЕРАЛЬНОЕ ГОСУДАРСТВЕННОЕ БЮДЖЕТНОЕ
ОБРАЗОВАТЕЛЬНОЕ УЧРЕЖДЕНИЕ ВЫСШЕГО ОБРАЗОВАНИЯ
«САРАТОВСКИЙ НАЦИОНАЛЬНЫЙ ИССЛЕДОВАТЕЛЬСКИЙ
ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ
ИМЕНИ Н. Г. ЧЕРНЫШЕВСКОГО»

Кафедра дискретной математики и
информационных технологий

**МОДЕЛИ ТЕОРИИ РАСПОЗНАВАНИЯ ОБРАЗОВ В
ЗАДАЧАХ ДИАГНОСТИКИ ЗАБОЛЕВАНИЙ ЖИВОТНЫХ**

АВТОРЕФЕРАТ БАКАЛАВРСКОЙ РАБОТЫ

студента 4 курса 421 группы
направления 09.03.01 — Информатика и вычислительная техника
факультета КНиИТ
Павлухина Максима Александровича

Научный руководитель

к.ф. -м.н., доц.

Ю. А. Бродская

Заведующий кафедрой

к.ф. -м.н., доц., зав. каф.

Л. Б. Тяпаев

Саратов 2018

ВВЕДЕНИЕ

В настоящее время такое понятие, как распознавание образов, получило широкое распространение во всяческих сферах человеческой деятельности — они используются сейчас повсеместно, в частности и в целях государственной важности.

Распознавание образов — новый шаг в развитии цифровой техники, алгоритмов и программирования. Благодаря ей можно решить множество проблем, с которыми сталкиваются программисты, а именно "решить задачу без алгоритма".

Сейчас эта сфера стабильно двигается и развивается, в частности в крупных компаниях, таких как Microsoft, Google и т.д. Соответственно, изучение и решение проблемы распознавания образов, которое рождает огромное количество приложений в различных сферах жизнедеятельности человека, чрезвычайно важно в современном мире.

Работа содержит в себе четыре главы:

1. Теория распознавания образов.
2. Основные подходы в теории распознавания образов.
3. Модели теории распознавания образов в задачах диагностики заболеваний животных.
4. Разработка приложения диагностирования заболеваний у животных.

Целью работы является разработка и реализация интеллектуальной системы диагностирования различного рода заболеваний животных. Акцент сделан на животных потому, что для людей такие системы уже разработаны, хоть и не в полном ключе, и действуют, есть конкурирующие друг с другом решения. Что же касается животных, то такого решения нет.

В ходе работы необходимо:

1. Изучить различные статьи по теории методов распознавания.
2. Разработать алгоритмы классификации с помощью различных методов распознавания в диагностике заболеваний животных.
3. Реализовать вышеперечисленные алгоритмы на языке программирования Python и сравнить их производительность.
4. Реализовать пользовательский интерфейс.

1 Основное содержание работы

Методы диагностики заболеваний у животных ничем не отличаются от методов диагностики заболеваний у людей. Меняются лишь величины основных показателей. Стоит отметить, что для людей реализовано достаточно большое количество различных диагностирующих интеллектуальных и более простых систем, эти системы широко применяются, а вот для животных таких решений нет.

Любые методы распознавания построены с помощью создания симуляции действий реального специалиста в той или иной области. Он производит изучение данных об объекте исследования и систематизирует эти данные — эту схему и нужно придерживаться в работе алгоритма.

Основными этапами являются:

1. Построение классификации. Этот этап требует некоторый уровень наших знаний об объекте, так как с помощью знаний и происходит группировка по принципу сходства и различия, сравнивая объекты между собой и эталонными объектами. Стоит отметить отличия в подходе к классификации человеком и машиной. Машина определяет вероятность принадлежности классу — человек же исходит из своей интуиции. Также отметим, что просто классифицировать, имея пару признаков, сложно — когда их сотни.
2. Классифицирование объектов. Человек классифицирует объекты на основе сравнительного анализа. При этом каждый из способов для знакомого с ним человека кажется простым. К этому стоит добавить тот факт, что классифицировать объект разные люди могут по-разному, что показывает отсутствие строгих правил в этой теории, однако все же можно выделить два подхода к классификации объектов:
 - а) Поиск более ближайшего эталонного объекта.
 - б) Формирование абстрактного понятия некоторой области пространства для каждого класса объекта.

2 Разработка приложения диагностирования заболеваний у животных

Была поставлена задача разработки приложения — интеллектуальной системы диагностирования заболеваний у животных с возможностью выбора метода распознавания из нескольких. После реализации системы, необходимо было произвести ее тестирование, исследовать ее поведение в различных случаях, найти наиболее производительный метод.

Одним из основных критериев для работы интеллектуальной системы диагностирования заболеваний животных является общедоступность: в физическом смысле и с точки зрения простоты использования. Самым идеальным вариантом для того, чтобы удовлетворить общий доступ к приложению с физической точки зрения — разработка веб-приложения. Удовлетворить простоту можно с помощью графического интерфейса и «понятных» для пользователя входных данных (например, фотография животного, которую он должен загрузить для обработки ее сервером). Простота не означает, что данной системой не сможет пользоваться квалифицированный специалист.

Для начала следует разбить задачу на некоторые составляющие:

1. Разработка метода простого перебора для диагностики заболеваний по результатам биохимических анализов.
2. Разработка наивного байесовского классификатора для диагностики заболеваний по результатам биохимических анализов.
3. Разработка нейронной сети для диагностики заболеваний по фотографии.

В свою очередь каждую из этих задач можно разбить на:

1. Изучение теории.
2. Разработка интерфейса.
3. Реализация метода и привязка его к созданному интерфейсу.

Следует отметить, что разработка каждого метода для любого вида животного требует дополнительных, достаточно больших медицинских исследований. В рамках данной работы классификация была проведена для заболеваний лошадей. Планируется дальнейшая работа над данным приложением (оно модульное и добавление новых функций, новых классификаций легко осуществимо). Кроме того, это приложение создано

заинтересовать широкий круг людей и выложено на GitHub (такую задачу никто не реализовывал в open-source-проектах).

2.1 Инициализация проекта в Django

Django — свободный фреймворк для веб-приложений на языке Python, использующий шаблон проектирования MVC [21, с. 616].

Обработчики URL в Django настраиваются явно при помощи регулярных выражений [14, с. 64–70].

Работа с базами данных представляет собой использование собственного ORM, в котором модель данных описывается классами Python, и по ней генерируется схема базы данных [15, с. 6–7]

Для начала создается виртуальное окружение с помощью virtualenv:

```
virtualenv -p python3 -no-site-packages venv
```

Активация окружения venv:

```
source venv/bin/activate
```

Внутри этого окружения необходимо произвести установку Python-модуль Django, а также команду:

```
pip freeze > requirements.txt
```

Она позволяет создать файл необходимых для проекта зависимостей.

Чтобы начать проект, необходимо выполнить команду:

```
django-admin startproject ADDT
```

Для приложения было использовано название ADDT (Animal Diseases Diagnostic Tool).

2.2 Создание базового интерфейса приложения

Базовый интерфейс веб-приложения представляет собой три блока:

1. Header.
2. Content.
3. Footer.

Все эти блоки могут быть описаны в общем для всех «оберточном» файле. Такой эффект достигается путем использования шаблонизатора Jinja 2 (генерации html-страниц с помощью шаблонов в случае с веб).

Bootstrap был использован для импорта таблиц стилей и генерации на их основе собственных.

Отметим, что нельзя забывать о настройке django-проекта: в файле settings.py необходимо указать:

```
'DIRS': [(os.path.join(BASE_DIR, 'addt/templates'))],
```

2.3 Создание формы для биохимического анализа

Для работы методов «применение метода простого перебора для диагностики заболеваний по результатам биохимических анализов» и «применение наивного байесовского классификатора для диагностики заболеваний по результатам биохимических анализов» необходимо было разработать форму для ввода данных биохимического теста.

Для этого сначала было необходимо выделить основные признаки, по которым в дальнейшем производилась классификация [18, с. 74–126]. Было произведено исследование, в ходе которого такие признаки были найдены:

1. АСПАРТАМИНОТРАНСФЕРАЗА (АСТ, АсАТ; AspАТ).
2. АЛАНИНАМИНОТРАНСФЕРАЗА (АЛТ, АлАТ; ALT).
3. КРЕАТИНФОСФОКИНАЗА (КФК, КК; СК).
4. ГАММА-ГЛУТАМИЛТРАНСФЕРАЗА (ГГТ; GGT).
5. КИСЛОТНОСТЬ (pH).
6. ХЛОР (Cl).
7. ТРИГЛИЦЕРИДЫ (TRIG).
8. ХОЛЕСТЕРИН (ТС).
9. АЛЬБУМИН (Alb).
10. КАЛЬЦИЙ ОБЩИЙ (Ca).
11. ЖЕЛЕЗО (Fe).

Воспользуемся модулем Forms фреймворка Django, чтобы получить необходимую форму.

2.4 Применение метода простого перебора для диагностики заболеваний по результатам биохимических анализов

Данный метод является самым примитивным и простым, но это дает ему огромный плюс в производительности. Данный метод «забирает» данные с заполненной пользователем формы для биохимических анализов, «видит», какое животное было выбрано, после чего производится сравнение по показателям. Когда сравнение окончено, выводится результат. Оптимизация данного метода была произведена путем выбора сначала наиболее важных

признаков, которые несут под собой более серьезные болезни.

Алгоритм представляет собой чтение биохимических данных, сравнение по референтным интервалам наиболее важных показателей, затем, если они в норме, проводится более глубокое сравнение по остальным показателям. На выходе имеем результат: больно ли животное, и какой болезнью, если больно.

На рисунке ниже можно увидеть общий результат (больно ли животное) и диагнозы, которые ставятся с помощью разработанного метода.



Рисунок 2.1 – Страница результата

2.5 Применение наивного байесовского классификатора для диагностики заболеваний по результатам биохимических анализов

Метод классификации, основанный на теореме Байеса, тоже достаточно прост, но он учитывает статистику, динамику и развитие заболеваний.

В данной работе был использован мультиномиальный байесовский классификатор (обобщающий биномиальное распределение на случай независимых испытаний случайного эксперимента с несколькими возможными исходами). Данный тип байесовского классификатора очень точен в случае наличия большого количества классов.

Алгоритм представляет собой чтение биохимических данных, подачу на вход мультиномиальному байесовскому классификатору. На выходе имеем результат: больно ли животное, и какой болезнью, если больно.

Страница результата имеет аналогичный рисунку 2.2 внешний вид.

2.6 Применение нейронной сети для диагностики заболеваний по фотографии

Данный метод классификации основан на нейронной сети, уже совсем непрост, но он более чутко учитывает статистику, динамику и развитие заболеваний.

В данной работе была использована сверточная нейронная сеть. Для ее построения была использована библиотека TensorFlow [17, с. 12]. Была использована Mobile Net 1.0 224 px, которая дает крайне хорошие результаты, даже при малой выборке [16, с. 6].

Алгоритм представляет собой чтение биохимических данных, подачу на вход нейронной сети. На выходе имеем результат: больно ли животное, и какой болезнью, если больно.

Страница результата имеет аналогичный рисунку 2.2 внешний вид с единственным добавлением, что выдаются не только наиболее вероятные болезни, но и остальные, с указанием их вероятности. Время работы составляет в среднем 0.150 мс — и оно достаточно высокое, чтобы назвать эту систему производительной.

Также была разработана форма для загрузки изображений.

ЗАКЛЮЧЕНИЕ

В данной работе были изучены и использованы в качестве решения различные модели теории распознавания образов.

В ходе работы:

1. Были изучены различные статьи по теории методов распознавания (в том числе на английском языке).
2. Были разработаны алгоритмы классификации с помощью различных методов распознавания (метод перебора с оптимизацией, наивный байесовский классификатор, нейронная сеть)
3. В качестве примера был выбран набор признаков из сферы ветеринарии.
4. Были реализованы вышеперечисленные алгоритмы на языке программирования Python.

В результате данного исследования можно сделать выводы:

1. Метод перебора с оптимизацией самый быстрый (в среднем его работа составляет 0.01 сек), однако он может выдать на входе несколько классов, поэтому окончательное решение принимает человек.
2. Наивный байесовский классификатор хорош тем, что достаточно быстродейственен (при количестве данных в 100000 экземпляров в среднем работа составляет 0.05 сек). Его существенным минусом является то, что он предполагает под собой независимость заболеваний, которые он определяет, но в действительности зависимость есть, причем достаточно существенная.
3. Нейронная сеть имеет под плюсами простоту в получении входных данных (в вышеперечисленных нужны результаты биохимических анализов, что само под собой требует наличия лаборатории, опытных сотрудников и времени). Для нейронной сети в данном случае достаточно лишь фотографии. К минусам относится быстродействие, которое обычно в среднем составляет 0.15 сек.

Самым перспективным в развитии является метод с применением нейронной сети. Система полностью готова к работе с любым, даже неподготовленным, пользователем.

СПИСОК ИСПОЛЬЗОВАННЫХ ИСТОЧНИКОВ

- 1 Лепский А.Е., Броневи́ч А.Г. Математические методы распознавания образов: Курс лекций/ А.Е. Лепский, А.Г. Броневи́ч. Таганрог: Изд-во ТТИ ЮФУ, 2009. 155 с.
- 2 Парамонов П. А. Методы, алгоритмы и устройства распознавания речи в ассоциативной осцилляционной среде: дис. ... канд. тех. наук: 05.13.05 / Павел Александрович Парамонов; науч. рук. И. В. Огнев. М., 2014 год. 139 с.
- 3 Открытые семантические технологии проектирования интеллектуальных систем Open Semantic Technologies for Intelligent Systems (OSTIS-2012): материалы II Междунар. научн.-техн. конф. (Минск, 16-18 февраля 2012 г.)/ редкол. : В. В. Голенков (отв. ред.) Минск : БГУИР, 2012. 548 с.
- 4 Вьюгин В. Математические основы теории машинного обучения и прогнозирования/ В. Вьюгин. М.: МЦМНО, 2013. 390 с.
- 5 Левченко В. Г., Достлев В. Г. Система на базе 3D технологий для автоматического выявления конструктивных дефектов технических объектов / В. Г. Левченко, Ю. С. Достлев// Информационные управляющие системы и компьютерный мониторинг. Донецк : ДонНТУ, 2012. 192 с.
- 6 Беркетов Г. А. и др. Решение задач прогнозирования состояния и управления жизненным циклом сложных технических комплексов методами распознавания образов/ Г. А. Беркетов. и др.// Статистика, экономика и информатика. Вестник УМО. 2014. №1. 143 с.
- 7 Котов К. А., Токарев В. Л. Определение вектора признаков в задаче идентификации по цифровому изображению/ К. А. Котов, В. Л. Токарев// Известия тульского государственного университета. Технические науки. Тула: ТулГУ, 2009. 199 с.
- 8 Кононюк А. Е. Общая теория распознавания Кн.1./ А. Е. Кононюк. Киев: «Освіта України», 2012. 586 с.
- 9 Гулевич С. П. и др. Методика селекции признаков классификации в задачах распознавания образов сложных пространственных объектов/

- С. П. Гулевич и др. // Наука и образование: электронное научно-техническое издание, 2012. 22 с.
- 10 Теоретические основы обработки и интерпретации геологогеофизических данных: метод. указ. для лабораторных занятий и самостоятельной работы магистров всех форм обучений по направлению 230400 «Информационные системы и технологии» по дисциплине «Теоретические основы обработки и интерпретации геологогеофизических данных»/ сост. С. К. Туренко, Ж. М. Колев. Тюмень: Издательский центр БИК ТюмГНГУ, 2013. 23 с.
 - 11 Кокаия М. П. Анализ базы данных траекторий автомобилей/ М. П. Кокаия. СПб, 2016. 34 с
 - 12 Django Documentation, Release 2.2.dev20180530015251: Django software foundation. 2018. 1880 с.
 - 13 Klimcikova lucia Virtual classroom in django web development/ Lucia Klimcikova. 2011. 60 с.
 - 14 Ravindran Arun Django design patterns and best practices/ Arun Ravindran. Packt Publishing, 2015. 222 с.
 - 15 Gendreau Thomas B. Using Python, Django and MySQL in a Database Course/ Thomas B. Gendreau. 2011. 15 с.
 - 16 Wang H. How can you get started with machine learning?/ H. Wang. 2017 год. 54 с.
 - 17 Doan E. Intro to TensorFlow and Machine Learning/ Doan E. 2017. 34 с.
 - 18 Пэворд Т., Пэворд М. Полный ветеринарный справочник по болезням лошадей/ Т. Пэворд, М. Пэворд. М.: Аквариум, 2005. 256 с.
 - 19 Лелевич С.В. Основы клинической биохимии: пособие для студентов медико-диагностического факультета / С.В. Лелевич и др. Гродно: ГрГМУ, 2013. 184 с.
 - 20 Егорова И. Э. Краткий курс биохимии. Часть I: учебное пособие для иностранных студентов / И. Э. Егорова и др. Иркутск: ИГМУ, 2013 год. 68 с.

- 21 Файцетт, Л. Fundamentals of Neural Networks: Architectures, Algorithms
And Applications: США: Pearson, 1993. 461 с.: ил.