***ФГБОУ ВО «Кубанский государственный аграрный университет имени И.Т. Трубилина»,***

***Российская Федерация***

***Потемин Илья ПИ2103***

**126 Теория информации, данные и знания. Лабораторная 6. работа 3.0.2 обработка текстов 2020-10-08**

Анализ изображений и их спектров в системе Aidos: Лабораторная работа №6

Резюме текста

1. Введение и организационные моменты:

Лектор начинает занятие 8 октября 2020 года (пятая пара, 15:35-17:05) для группы ИТ-2002 по дисциплине "Теория информации, данные, знания".

2. Уточнение темы занятия: Переход от обработки текстов к изображениям:

Изначально планировалось начать изучение возможностей интеллектуальной обработки текстов (Лабораторная работа 3.02). Однако выясняется, что эта тема уже была пройдена. Лектор корректирует план и переходит к теме обработки изображений в системе Aidos (Лабораторная работа №6), начиная с анализа изображений флагов.

3. Подготовка данных для обработки изображений (Конвертация и изменение размера):

Демонстрируется работа с набором из 30 изображений флагов разных стран.

Подчеркивается, что система Aidos работает с форматами JPEG и BMP, а исходные файлы были в PNG.

Показан процесс конвертации файлов из PNG в JPEG с помощью программы ACDSee.

Объясняется необходимость изменения размера изображений (ресайз) до приемлемого для Aidos (например, 200 пикселей по ширине), так как система имеет ограничения на визуализацию очень больших изображений (до ~450 пикселей) и их обработка занимает больше времени. Ресайз также выполняется пакетно в ACDSee.

4. Работа с системой Aidos: Загрузка и обработка изображений (Интерфейс 2.3.2.3):

Подготовленные JPEG-файлы копируются в папку inpdata системы Aidos.

В системе Aidos удаляются предыдущие приложения.

Используется подсистема формализации предметной области.

Создается база данных изображений, и в нее загружаются подготовленные файлы флагов из папки inpdata.

Запускается процесс создания файла inpdata (табличного представления) для программного интерфейса 2.3.2.3, который предназначен для ввода изображений по пикселям и спектру.

5. Функционал системы Aidos: Подсистемы и интерфейсы:

Кратко описывается подсистема формализации предметной области, включающая работу с классификационными/описательными шкалами и обучающими выборками.

Объясняется понятие API (программных интерфейсов) в Aidos как программ, преобразующих данные из одного стандарта в другой.

Перечисляются различные типы интерфейсов для ввода данных: текстовых, табличных, графических (по контурам или по пикселям/спектру).

6. Создание обучающей выборки и синтез моделей:

Запускается интерфейс 2.3.2.3 для формализации данных – создания шкал градаций и обучающей выборки на основе загруженных изображений флагов.

Демонстрируется просмотр созданных классификационных шкал (названия флагов) и описательных шкал (спектральные диапазоны, представленные как коды пикселей).

Выполняется синтез и верификация моделей распознавания на графическом процессоре.

7. Оценка достоверности моделей:

Анализируется достоверность созданных моделей. Модель F5 с интегральным критерием "Сумма знаний" показывает результат 0.936, что лектор характеризует как "абсолютно идеальную модель", практически безошибочную.

Эта наиболее достоверная модель выбирается как текущая для дальнейшей работы.

8. Решение задачи идентификации:

Запускается задача идентификации (распознавания) на основе выбранной модели F5.

9. Анализ результатов: Спектры, дендрограммы, матрицы расстояний:

Демонстрируются результаты идентификации.

Показывается возможность просмотра спектров изображений (распределение пикселей по спектральным диапазонам).

Объясняется использование дендрограмм и матриц расстояний для визуализации сходства и различий между объектами (флагами) по их спектрам.

Упоминается возможность просмотра информационных портретов объектов (наиболее характерные и нехарактерные признаки/цвета для каждого флага).

10. Теоретические аспекты и возможности системы:

Кратко затрагивается теоретическая основа системы Aidos как метода, разлагающего функцию, описывающую объект, в ряд по функциям классов, что является обобщением рядов Фурье, Тейлора и др.

11. Задания для студентов и регистрация в научных сетях (РИНЦ, ResearchGate):

Подчеркивается необходимость выполнения студентами собственного проекта (приложения) в системе Aidos.

Обсуждается важность регистрации в РИНЦ (Science Index) и ResearchGate для размещения своих работ.

Поднимается проблема отсутствия у студентов корпоративной почты (@edu.kubsau.ru), необходимой для ResearchGate. Предлагается workaround: создать сайт виртуальной научной организации на бесплатном хостинге (например, PyHost), получить для него домен и почту на этом домене, что позволит зарегистрироваться.

12. Планы на следующее занятие:

Рассматриваются два варианта: либо разобрать реальную задачу (работа 3.04 - оценка автомобилей), либо предоставить время студентам для работы над их индивидуальными проектами и консультировать их.

Детальная расшифровка текста

1. Введение и организационные моменты

Поэтому еще раз скажу: здравствуйте, ребята.

Здравствуйте.

Здравствуйте, здравствуйте.

Сегодня 8 октября 2020 года. Пятая пара, которая идет с 15:35 по 17:05.

Лабораторная работа номер шесть у группы ИТ-2002 по дисциплине "Теория информации, данные, знания".

2. Уточнение темы занятия: Переход от обработки текстов к изображениям

И у нас по плану на этой лабораторной работе мы начинаем изучение возможностей, которые имеются в настоящее время, интеллектуальной обработки текстов.

Это лабораторная работа 3.02.

То есть мы изучаем с вами лабораторную работу 3.02, ребята.

Сейчас. Если что-то не так, вы мне сразу говорите. Мы с вами еще не изучали эту работу, да? Обработку текстов.

Мы на последнем занятии остановились на том, что с вашей группой, о том, что...

Исследование, исследование предметной области, развитый алгоритм принятия решений, исследование моделирования предметной области.

Вот это было последнее у нас, что мы с вами рассмотрели.

Или мы уже, или мы уже изучали лабораторную работу 3.02, обработку текстов, ребят?

Изучали.

Изучали.

Вот.

Тогда мы сейчас изучаем обработку изображений в системе Aidos.

Возможность обработки изображений.

Надо сказать, обработка изображений.

То есть мы табличные данные прошли.

Текстовые данные, интеллектуальная обработка текстовых данных прошли.

Теперь мы изучаем обработку изображений.

Интеллектуальная обработка изображений.

Что-то я несколько упустил, как-то сбился немножко с того, что мы должны изучать. Но, видите, слава богу, разобрались.

3. Подготовка данных для обработки изображений (Конвертация и изменение размера)

Ребят, значит, мы сейчас используем небольшой пример, состоящий из тридцати изображений, для того, чтобы освоить вкратце некоторые возможности посмотреть.

Значит, сразу вам даю ссылочку на исходные данные по этой задаче. Эти данные, они не на Kaggle, не на каком-то сайте для, на котором расположены данные для машинного обучения, а просто сайт. Я вчера просто поискал немножко, нашел изображения флагов разных стран.

И хочу вам сейчас показать, как можно обрабатывать эти изображения этих флагов, прямо графические объекты.

Ну, прежде всего, хочу вам вот что сказать, что исходный архив, файлы были PNG. Видите, да? Система Aidos работает с форматами JPEG и BMP. Поэтому я эти файлы скопировал в папочку и преобразовал их в JPEG. Я это сделал с помощью ACDSee программы. Вы знаете эту программу, нет? Знакомы с ней? ACDSee.

А вы сейчас демонстрируете экран?

Нет. Нет, сейчас я открою доступ, буду демонстрировать. Пока что я э-э просто вам рассказывал.

Значит, вот смотрите, значит, я сейчас вам покажу PNG архив. Ссылочку я послал вам на архив этот.

Значит, это большие файлы высокого разрешения. Ну не все, но многие там довольно большого, до 2000 пикселей. Я их скопировал в папочку в отдельную и с помощью программы ACDSee преобразовал их в JPEG.

Ну давайте я вам покажу, как это выглядит, чтобы вы представляли. Вот я беру, допустим, эти файлы, копирую вот сюда из архива.

Вот. И открываю в программе ACDSee.

Вот. Вот они открываются. И мы видим, что у них разрешение 2560 на 1280. Ну такие приличные. Ну есть поменьше, но, по-моему, мало. В основном вот, да, они вот такого все разрешения.

Вот. Теперь смотрите, значит, нам, это PNG файлы. Я их отмечаю все эти файлы, выделяю их все. И э-э Tools, э-э формат конверт в JPEG с удалением исходных файлов.

Вот. Раз, пошло.

Значит, программа ACDSee для таких целей является, ну, на мой взгляд, вообще идеальной. Прекрасно она работает, очень хорошие у неё функции нужные. Э-э пакетное переименование, пакетное преобразование из одного формата представления графических данных в другой, из одного стандарта в другой. Пакетное изменение размеров. Вот сейчас Resize. Batch Resize.

Вот. Всё. Теперь, сейчас это у нас JPEG файлы теперь. Но они все еще высокого разрешения. Я, значит, беру э-э теперь... Кстати, вот этот Resize, изменение размеров, он работает не со всеми типами файлов. Вот JPEG он открывается эта возможность и вращение, и э-э изменение их выдержки, скажем так, подбор выдержки. Вот. А с PNG файлами это закрытые возможности. Значит, я вот открываю, видите, в JPEG файлах уже указываю 200 пикселей по ширине. Ну, конечно, надо их все блоком выделить.

Вот она их все масштабирует. Для чего? Дело в том, что система Aidos, она, когда обрабатывает изображения вот в этих режимах, которые сейчас я буду показывать, нежелательно ей давать э-э очень большие изображения по пикселям. Почему? Она дольше будет обрабатывать и визуализировать в этих экранных формах, которые в этом режиме, она может там до 450 пикселей, к примеру, примерно. Там же не одно изображение, там же еще и спектры этих изображений, другая информация. Получается, что изображение очень большого размера неудобно обрабатывать. Поэтому, значит, вот э-э я масштабирую их, используя ACDSee. ACDSee. Вы с этой программой знакомы, ребята? Скажите, вообще есть у вас она, видели её?

А?

Лично нет.

Есть, да? Ну, удобная.

Вот есть возможность пакетного переименования файлов. Очень хорошая тоже функция, позволяющая пронумеровать файлы, сделать название у него из какого-то там, какой-то текст, и потом э-э номер файла. Тоже в системе Aidos используется такая возможность.

4. Работа с системой Aidos: Загрузка и обработка изображений (Интерфейс 2.3.2.3)

Значит, вот мы видим файлы, уже теперь у нас графические, э-э JPEG представлены в формате эти графические файлы. И мы можем теперь взять небольшое количество этих файлов, обработать в системе.

Значит, э-э я могу вам сказать, что для этого надо перенести эти файлы в папочку inpdata. Эта папочка для входных данных.

Вот, то есть те файлы, которые мы будем обрабатывать, перенесём.

Я, кстати, все их обрабатывал, но сейчас вам хочу сказать, что на лабораторной работе, для ускорения процессов, мы возьмём не все. И на смысл это особо не влияет. Вот, то есть вы увидите всё то же самое, только быстрее.

Вот. И, значит, что мы делаем? Заходим в систему.

Вот, удаляем все приложения, какие там были.

Да, ребята, значит, перекличка. Вы все присутствуете. И больше того, я тоже присутствую.

Вот. И идёт запись у нас.

Вот, то есть мы удаляем э-э все приложения, которые были в системе. И заходим э-э в подсистему формализации предметной области.

5. Функционал системы Aidos: Подсистемы и интерфейсы

И в этой подсистеме у нас есть что? Возможность классификационные шкалы и градации посмотреть, описательные шкалы и градации. И вот обучающие выборки осуществить, а также посмотреть на базы событий.

Вот. При этом вот обучающие выборки возможны в разных вариантах. Это ручной ввод обучающей выборки и корректировка, а также программные интерфейсы ввода данных из внешних источников данных, управление обучающей выборкой и декодирование сочетания признаков обучающей выборки. Некоторые из этих режимов, они еще пока не разработаны в этой версии новой. А в старой версии они были. Я оттуда перенёс это меню фактически.

Вот. Значит, смотрим теперь на программные интерфейсы. Что такое программные интерфейсы? API, Automatic Program Interface. Э-э это автоматизированные программные интерфейсы по-русски. Это программы или режимы системы, ну то есть тоже, по сути, программы в составе системы, которые обеспечивают преобразование данных из одного стандарта представления данных в другой стандарт представления данных.

Значит, мы видели, как мы использовали э-э графические, то есть текстовые стандарты представления данных э-э в лабораторной работе, табличные стандарты. Табличные, значит, мы не все эти интерфейсы изучали, но я вам на примере интерфейса 2.3.2.2 показывал, что можно обрабатывать и числа, и числовую, и текстовую информацию, решать задачи по математической лингвистике.

А теперь, значит, э-э есть еще интерфейс ввода файлов текстовых, прямо не фрагментов этих файлов, там абзацев, которые я вам показывал, как в экселевский файл легко вставляются, а прямо текстовых файлов любого размера. Книжки я обрабатывал, смотрел, как получается. Нормально получается, интересно. Вот. А также импорт данных из табличных э-э э-э стандартов, табличных, но не такого, как, не такого стандарта, как интерфейс 2.3.2.2.

А также возможно ввод, возможен ввод данных, ребята, в систему Aidos графических. При этом э-э может обрабатываться изображение по контурам, то есть только контуры изображения вводятся в систему, а также может обрабатываться целиком всё изображение по пикселям. Тогда изучается, есть возможность изучения спектра этих изображений.

А также есть интерфейс ввода текстовых э-э и числовых рядов, символьных, цифровых и э-э э-э то есть символы, в том числе цифры. Вот. И некоторые служебные режимы, которые бывают нужны в этой связи. Вот, это три вот этих режима. А также есть тут программные интерфейсы для чемпионата по системам искусственного интеллекта, Открытый чемпионат России в 2017 году был. Я про него говорил, я там участие принял в трёх номинациях, в двух первое место занял, в одной третье место. И есть очень развитый интерфейс решения решения задач, связанных с землетрясениями. Это прямо целая подсистема. По ней книги есть, статьи, патенты. Не патенты, да, тоже есть.

6. Создание обучающей выборки и синтез моделей

Но сейчас мы э-э используем программный интерфейс для ввода изображений по всем пикселям и спектру.

Значит, здесь у нас есть два варианта. Один вариант мы осуществляем формализацию предметной области, и второй вариант - мы создаём распознаваемую выборку. Когда мы формализация предметной области, что под этим понимается? Я вам уже про это рассказывал. Под этим понимается создание классификационных и описательных шкал и градаций и кодирование исходных данных с их использованием. И формирование в результате обучающей выборки. А потом, значит, на следующем этапе мы будем эту обучающую выборку использовать для синтеза модели, решения задач.

Но сейчас мы осуществляем формализацию, то есть мы создаём шкалы и градации и обучающую выборку. Извините.

Для этого мы создаём базу данных изображений с нуля. Там могла быть база изображений, в которую уже включены изображения. Потом загружаем эти изображения из папочки inpdata в базу э-э изображений, Image Database.

Вот они загружаются. Если бы они были больше по площади, э-э по числу пикселей, они бы загружались дольше. Если есть желание, можно посмотреть, какие там изображения в базе.

Ну, обычно такого желания нет. Просто я сразу пересоздаю и создаю, заношу туда изображения. А потом, обратите внимание, создать файл inpdata для программного интерфейса 2.3.2.3. Это табличный программный интерфейс, значит, который позволяет, который я использовал для ввода изображений. Вот, стандартный интерфейс, причём, системы. Значит, здесь все параметры стоят по умолчанию, э-э потому что они были созданы вот в этом режиме до этого. Но, в принципе, можно что-то менять, если есть желание.

Тут есть разные варианты: число градаций, э-э число спектральных диапазонов в спектре, э-э только по спектру, только по пикселям, по пикселям и по спектру, как кодировать э-э истинно чёрный цвет. Вот. Как кодировать истинно белый цвет, учитывать ли фон. Ну, в общем, всякие такие вот дела тут. Всё это я откладку ввёл когда-то в своё время. Но потом вот сегодня я разместил новую версию от 8.10.2020, в общем, ну, на сайте, э-э в общем доступе, обновление разместил на этой версии. Почему? Потому что оказалось, что я кое-что там покорочил. Ну, в процессе совершенствования некоторые возможности системы были повреждены. Вот сегодня я их, вчера, сегодня я их восстановил и разместил новую версию, которая работает так, как раньше. Да, кстати, сейчас я вам э-э, пока она там будет думать, я вам покажу... Да, кстати, обработка графики – это дело такое, э-э затратное в вычислительном отношении.

А что там за проблема была?

Проблема была в том, что э-э когда э-э спектральный анализ проводится, там у меня э-э в базах данных было м-м м-м название, наименование градаций полей были такое: спектр интервал, спектральный интервал, ну спектральный диапазон. И потом сам диапазон указан. Так я взял, умудрился каким-то образом кое-где, не везде, а кое-где поменять его, написано маленькими буквами почему-то он стал. Не большими буквами спектр интервал, а маленькими. В результате в некоторых местах перестал работать поиск, и перестало отображаться форма, формы, э-э экранные формы, выходные формы перестали отображаться в таком виде, в каком они должны отображаться при исследовании спектров. Сейчас я вам сошлюсь на статью, вернее, даже не на одну статью, а на несколько статей. Ну я просто взял большими буквами опять сделал. Но пришлось, конечно, там по программе поиском. Вот, автоматизировано я уже не стал делать, потому что автоматизированная замена, вот она, видимо, и привела к этому.

Значит, э-э смотрим. Ну то есть просто я покорочил программу немножко в процессе совершенствования. Лучшее - враг хорошего, знаете ж.

Это страничка, где у нас скачивается система. Поэтому вам желательно систему обновить.

Значит, вот ссылочка на статьи по анализу изображений, которые в системе. То есть довольно много статей. Вот есть. Сейчас я в чат отправлю эту ссылочку. При желании, значит, можете посмотреть.

И ссылку на обновление сейчас пришлю системы.

Ну она обновляется, если StarTedos запускать. Но я так понимаю, что иногда на некоторых компьютерах он выдаёт ошибку в десятой, под Windows 10.

Вот. Ну, вроде вот закончился процесс. То есть, видите, здесь прямо написано для пользователя, что подготовлена база данных для программного интерфейса 2.3.2.3 (2.3.2.3). Поэтому мы в этот интерфейс мы перейдём. То есть здесь уже у нас э-э все пункты выполнены э-э программного интерфейса ввода данных. И э-э он как бы является прединтерфейсом, если честно. То есть он подготовил данные для стандартного программного интерфейса. То есть графические данные были преобразованы таким образом, что можно теперь использовать стандартный программный интерфейс. Вот как он выглядит. Он более такой упрощённый, чем 2.3.2.2. Меньше тут возможностей. Но э-э 2.3.2, 2.3.2.2 у него было что? У него миллион строк было и до 16.000 колонок. А в этом наоборот. Э-э, ну при этом колонки мы использовали часть колонок как классификационные шкалы, а часть как описательные шкалы. А здесь наоборот, у нас строчки являются описательными, классификационными и описательными шкалами, а колонки являются наблюдениями. То есть этот режим позволяет таким образом э-э вводить э-э данные табличные, в которых до миллиона, грубо говоря, э-э классификационных и описательных шкал и градаций, суммарно. Ну это безумно большое количество, никогда столько не нужно. Но почему я его сделал? Потому что иногда было нужно больше 16.000, понимаете? Причём я его сделал тогда, когда ещё было не 16.000, а 256 э-э всего лишь колонок было в экселевском файле. Поэтому возникал вопрос, что как-то маловато шкал. И я придумал, что можно использовать строки в качестве шкал, а колонки в качестве наблюдений. И тогда э-э дело пошло.

Вот. Ну здесь вот мы видим, что у нас э-э номер первой строки классификационной, номер последней строки классификационной. То есть у нас здесь не колонки, не диапазоны колонок, а диапазоны строк указываются. И указывается диапазон строк описательных шкал со второй по тридцать шестую, потому что это это э-э спектральные диапазоны. Вот. Ну и, в общем, здесь вот своя специфика есть определённая. Ну, в общем, э-э то, чего там не получилось, получилось здесь. И вводятся та информация, которая была сформирована предварительным программным интерфейсом, прединтерфейсом вводится в систему. Теперь мы можем посмотреть, как и раньше, что у нас представляют собой классификационные шкалы. Это у нас эти вот 30 флажков. Вот. И что представляют собой описательные шкалы? Вот это вот не работало, ребята. То есть здесь у нас спектральный интервал, видите, вот написано, спектральные диапазоны. И заданы э-э 35 спектральных диапазонов. Вот это вот перестало визуализироваться этот спектральный диапазон, понимаете? Почему я и э-э немножко поработал вчера.

Ну сейчас это визуализируется, и в других выходных формах та же самая ситуация. Причём, почему-то кое-где он визуализировался, остался работающим. Кое-где. А кое-где вот перестал. Но теперь везде работает. Здесь вот у нас указаны э-э эти RGB э-э представления цвета. Green Blue. Указаны яркости этих лучей RGB: Red, Green и Blue. Для этого цвета, для этого цвета и так далее. Ну здесь не синий, а просто здесь курсор стоит на этом месте. Ну я вот сюда его поставлю, тогда будет не так заметно. Вот. И всё, вот идёт он. Вот. Дальше у нас есть обучающая выборка, которая на вид имеет безумный, в общем-то. В каком смысле? Ну, здесь вот у нас вверху сами эти графические объекты, а здесь, ребята, коды пикселей в соответствии со справочником э-э спектральных интервалов.

Понимаете, да, что это такое? Это пиксели. Вот каждый код - это это код пикселя в соответствии с код спектрального диапазона точнее, к которому относится пиксель. А это вот э-э код класса. Вот.

И мы можем на сами эти изображения глянуть вот здесь. Вот, как они выглядят. Вот они. Это те, которые сейчас мы будем отслеживать.

7. Оценка достоверности моделей

Значит, э-э количество, ну я решал задачи такие с графическими объектами, где их сотни. Вот, но могу вам сказать, что кластерный анализ, он э-э ещё можно посчитать дендрограммы, когда их этих вот объектов, ну, скажем, около 300. Вот 240, там 250 ещё. 4К дендрограмма визуализируется, ещё понятно, что там такое. А когда их там, скажем, 500, тогда с этим могут быть вопросы, то есть э-э просто по размерности э-э форм выходных.

Ну вот я сейчас вам покажу дендрограмму для большого количества объектов графических. Это она э-э 4К по э-э высоте. По ширине поменьше, там не требуется. А по высоте здесь вот уже 4К используется. Видите? Здесь показано, какие флаги, какие государства сходны друг с другом. Ну это двухбуквенное обозначение страны. Вы, наверное, знакомы с такими обозначениями. Они везде используются. То есть некоторые флаги вообще очень похожи, видите? По цветовому решению, я бы сказал так.

Вот. Ну и матрица расстояний тоже 4К, видите, как выглядит. Ну можно было пошире сделать, тогда бы здесь вот эти э-э формулы кластерные влезли бы, которые внизу. Они здесь просто не поместились, длинные. Ну я не стал там из-за этого расстраиваться особенно. Вот. Ну это вот полная как бы вариант этого приложения.

Вот. И смотрим мы дальше. Э-э собственно, что мы сделали? Обучающие, классификационные описательные шкалы градации сделали. И теперь э-э осуществляем синтез и верификацию модели. Значит, всё делаем на графическом процессоре теперь уже. Я вам упоминал, что там была неточность в расчёте модели PRC2. Я её исправил эту неточность. Вот. Ну теперь мы, значит, э-э её исправил не путём модификации программы на OpenGL, а путём того, что после работы этой программы я просто взял и исправил э-э то, что там эта программа она не совсем правильно делала. Я это сделал в базовой системе, просто исправил это.

Вот. Смотрим, что у нас получилось с точки зрения достоверности модели.

Вот, модели получились у нас 0,936 по критерию L2, который я вам объяснял, что это нечёткая мультиклассовая обобщение критерия Вальдсбергена, инвариантно относительно объёма выборки. 0,936. Знаете, что это значит, ребят? 0,936? Это абсолютно идеальная модель, практически совершенно безошибочная. Вот, сейчас мы посмотрим на решение задачи в этой модели. Ну здесь это не очень понятно.

Ну здесь вот более понятно. Значит, мы видим, что ложных решений положительных нет. Ложных решений отрицательных нет. Вот. А, нет, вернее так, они-то есть, но э-э мы видим э-э закономерность какую, что начиная с 25% сходства резко увеличивается число истинных решений, и э-э начиная с 35% ложных решений нет. Вот. А решений отрицательных ложных вообще нет. Ну то есть модель F5 с интегральным критерием Сумма знаний является идеальной.

8. Решение задачи идентификации

Значит, вводим, значит, сейчас мы использовали не лабораторную работу для изучения, а своё приложение сделали. Это приложение вполне обоснованно назвать э-э спектральный э-э синтез и анализ изображений флагов. Ну вот так можно назвать. А ещё есть сценарный СК-анализ. Сейчас несколько статей написал в сентябрьский номер, э-э в октябрьский номер. И сейчас я вам дам ссылочку на будущие эти статьи. Да, ребят, вы обновите систему. Для этого надо из системы выйти, скачать этот файл обновлений и развернуть его в папочке с системой. Выйти из системы для чего? Чтобы исполняемый модуль можно было заменить на тот, который в обновлении.

Вводим, значит, э-э, сейчас мы её, э-э, то есть мы, э-э, выбрали наиболее достоверную модель, делаем её текущей. Делаем её текущей. Это для того, чтобы решить задачу идентификации. И решаем задачу идентификации в этой модели.

9. Анализ результатов: Спектры, дендрограммы, матрицы расстояний

Всё. Смотрим результаты решения задачи идентификации. Ну вы уже с этим знакомы. И видим, что у нас э-э с тем интегральным критерием, которым у нас э-э задача решается более достоверно – это сумма знаний. Я вам рассказывал про эти интегральные критерии на прошлых занятиях, объяснял всё это. Поэтому сейчас уже не объясняю так подробно.

Вот вы здесь видите, что, в общем, э-э как... Есть э-э и истинные решения, и ложные. Вот. Но, м-м, как-то так, в общем, она, в принципе, определяет, какой флаг.

Вот. И дальше мы смотрим, каким образом она это делает. Значит, да, во-первых, давайте сейчас чуть-чуть исследуем предметную область. Значит, возьмём и посмотрим, насколько эти флаги сходны друг с другом по их спектрам.

Вот. И посмотрим, что у нас получилось.

Вот она определила, какие из них сходны друг с другом по тем изображениям, которые там на них есть, ну по их, вернее так, не по самим изображениям, а по цветовым решениям, скорее.

Вот.

10. Теоретические аспекты и возможности системы

Значит, э-э я вам показывал информационный портрет. Это информация о том, какой цвет насколько характерен для того или иного флага. Вот, допустим, для этого флага характерны вот эти цвета. Вот для этого характерны э-э вот эти. А эти вот не характерны. Вот. То есть мы для каждого флага знаем, какие цвета характерны, а какие не характерны из того спектрального диапазона, который там был у нас.

И э-э с мы можем посмотреть на круговую диаграмму креативную, э-э которая э-э визуализирует э-э то, насколько сходны различные цвета по э-э по смыслу.

Значит, изображение, кстати, могут относиться к классам. Здесь я не показал. А здесь каждое изображение представляет собой класс. А мы можем эти изображения группировать э-э, ну, скажем так, несколько изображений использовать для формирования образа класса.

Вот. Что-то я в этом примерчике простеньком этого не показываю.

Вот. И здесь мы можем э-э вроде модель и посмотреть на круговую диаграмму. И в ней э-э показано, какие цвета сходны по смыслу, а какие отличаются, и на сколько. И вот эти вот э-э надписи, они сделаны на фоне, соответствующем этому цвету, который соответствует одному спектральному диапазону. Тоже можно, конечно, здесь уменьшить, увеличить. Вот. И посмотреть, что получится.

Ну здесь вот мы вверху видим, что это конструкт признака, спектральный интервал, первый. Это его цвет этого спектрального интервала и приложение Спектральный анализ изображений флагов.

И э-э дальше я вам покажу интересную вещь. Здесь вот у нас есть АСК-анализ изображения по пикселям, спектрам и контурам. Значит, здесь есть такие учебные режимы для того, чтобы э-э в учебном плане осваивать это всё. Ну, они не только учебные, ну, в общем, такие вот. А здесь вот есть э-э и этот выход на этот интерфейс ввода изображений, а также на формирование, ну, в общем, все те же самые вещи, что я показывал оттуда из главного меню, здесь есть внизу э-э в графическом этом интерфейсе этой системы. Здесь всё есть. Вот мы сейчас смотрим на спектры этих изображений. То есть она выводит изображение и то, каких пикселей сколько, какого спектрального диапазона. Здесь вот они внизу эти спектральные диапазоны в качестве шкалы X, а здесь э-э в процентах от всех э-э пикселей, сколько пикселей такого диапазона представлено.

Вот. И так вот мы видим, что они отличаются друг от друга.

Значит, э-э что здесь интересного, ребят? Значит, система выявляет спектры э-э изображений, э-э а потом э-э может формировать обобщённые образы на основе спектров изображений, которые относятся к одному классу. Ну, в частном случае этими классами являются сами э-э эти объекты обучающей выборки. То есть можно создать образ класса на основе одного объекта, а можно на основе нескольких. Но сейчас я вам показываю пример, где образы классов созданы на основе одного объекта.

11. Задания для студентов и регистрация в научных сетях (РИНЦ, ResearchGate)

Значит, э-э я вам рассказывал, что есть э-э спектральный анализ, да? А есть анализ контуров изображений. Значит, вот задача интересная, ребята. Значит, э-э нет, Петрович Трошин мне задачу мне предложил. Он говорит: "Так, вот смотрите, Евгений Вениаминович, вот есть вот листья э-э с разных винограда разных сортов. И э-э у них разная форма, э-э потому что сорта разные. Но, более того, даже на листе винограда одного и того же сорта есть листья разной формы. Как это вообще понимать? Можно ли по форме листа определить сорт винограда?" Это задача, которую он нам сформулировал. Я ему выразил такую мысль, что, ну, он сказал, что есть там адаптивность, есть влияние среды. То есть форма листа определяется двумя факторами: это геномом растения, генетически определена. Фактически геном одинаковый во всех листьях одного куста. Поэтому, значит, форма должна быть одинаковая, если она обусловлена была бы чисто геномом. Но она обусловлена не только геномом, но и воздействием окружающей среды. Поэтому она э-э имеет некую вариабельность. То есть фактически можно считать, что форма листа представляет собой сумму истинной формы листа сорта и шума какого-то, который обусловлен воздействием окружающей среды. Он говорит: "Можно ли выделить истинную форму листа, подавить шум и выделить истинную форму листа?" Ну он в этих терминах не говорил. Он просто сказал, что можно ли по форме листа сорт определить. Я говорю: "Можно попытаться". Значит, вот в системе Aidos есть интерфейс, позволяющий вводить контуры, контуры изображений. И получается в полярной системе координат э-э шкалы, градации. Здесь градации - это вот э-э эти полосочки цветные, а шкалы - это вот э-э сектора. Можно определить, э-э какие признаки наиболее характерны, наиболее не характерны для того или иного э-э для листов того или иного сорта. И выделить э-э те особенности этих листьев, которые наиболее характерны именно для данного сорта. Это было сделано. Вот прямо режимы соответствующие разработаны. То есть здесь показано только наиболее характерное значение э-э то есть наиболее характерные признаки для этого сорта. И, в общем, э-э задача решается идентификация э-э растения по форме листа, то есть сорта растения по форме листа. Это очень интересная задача. Значит, фактически э-э что можно сделать? Допустим, где-нибудь там в горах Абхазии там лазишь, смотришь, какой-то кустик растёт э-э дикого винограда. Берёшь его листики, привозишь сюда. И можно исследовать э-э этот сорт винограда, который там в горах Кавказа, он откуда здесь взялся? Либо это местный сорт, либо это э-э какой-то греческий сорт. Вот все такие задачи вот мы их решали с помощью этой технологии. И получилось, что у нас э-э виноград, который вот здесь у нас в Краснодарском крае, что это в основном греческого происхождения виноград. Даже в горах.

Значит, я должен вам показать статью интересную про это. Ну я вам ссылочку послал на эти статьи, но просто одну из них не статьи, а пособие покажу, где показана в развитой форме, как эта технология применяется. И сейчас ещё статью покажу одну. Смотрим. Ну, знакомая, да, картина? Значит, здесь о чём речь идёт? Значит, э-э о том, что э-э есть растения, у которых листья э-э разной степени повреждённости. И система измеряет степень повреждённости этих листьев, исследуя эти изображения, их спектр. Вот эти примеры этих. Допустим, от нуля до 80% повреждений. Вот эти 35 диапазонов. Я делал и больше, и меньше, но потом пришёл к выводу, что нет смысла делать больше, меньше нежелательно. Вот. И мы здесь видим, что вот эти спектры этих изображений этих листьев. А потом было много листьев, ребята, много, ну, скажем, там, я сейчас не помню, там 20 штук или 30, которые относились к одному уровню повреждения. Вот здесь видите, тире стоит, вот 0,2 там. Это вот у нас класс, ячмень 0,5% - это класс. А это вторая реализация э-э объекта, который относится к этому классу. А это первая реализация объекта, который к этому классу. И система формировала обобщённый образ класса. Это третий реализация, четвёртая реализация, всё к одному и тому же классу относится. А это уже другой класс, 10%. Он более коричневый. И вот спектры их сформированы. А потом дальше, смотрите, что происходит. Вот она закончила формирование спектров. А теперь, смотрите, что интересно. Значит, она может сформировать, поскольку образы классов сформированы, то она может вывести образы этих классов нужной модели. Вот. Что это, что это такое, ребята? Значит, если какой-то цвет имеет какую-то вероятность встречи безусловную по всей выборке, видите, модель PRC1, условные и безусловные вероятности встречи. Здесь это описывается внизу. Я вам это рассказывал, когда про модели рассказывал. То есть я рассказывал, что э-э сравнивается условное процентное распределение с безусловным. И сравниваются относительные частоты. То есть, ну, грубо говоря, так, вероятность встречи какого-то цвета у объектов какого-то класса сравнивается с безусловной вероятностью встречи этого цвета по всей выборке. И э-э формируется обобщённый образ уже в модели системно-когнитивной модели, там F1, F3, F5, это уже системно-когнитивные модели. F5 - это коэффициент возврата ROI э-э ну это они все похожи на Х квадрат. Я вам показывал это всё, объяснял. Вот. И что мы видим? Мы видим, что некоторые цвета более характерны для данного класса, чем в среднем. То есть они чаще встречаются у объектов данного класса, чем в среднем по всей выборке. А некоторые реже встречаются, чем у объектов всей выборки. Они встречаются, но они не характерны для этого класса. Ну это примерно как вот признак этот длинные волосы. Он встречается у ребят, но не характерен для них. И он будет вот здесь вниз идти. То есть он будет нести информацию отрицательно о принадлежности к этому классу. А это вот характерно. А это вот просто у нас э-э э-э среднее, вот это пунктиром - это среднее показано, э-э безусловная вероятность встречи этих цветов у объектов выборки. А это э-э спектр этого объекта, этого класса, вернее, спектр этого класса. А это спектр уже э-э именно показано сравнение спектра класса со средним. Вот. То есть здесь показано, какие цвета характерны, а какие не характерны. Вот. И для объектов этого класса. Это позволяет решать такую задачу. Если у нас э-э есть некий объект, мы его изображение вносим в систему стандартным интерфейсе, и он сравнивается с обобщёнными образами изображений. И фактически по его спектру определяется, какому классу он относится. То есть это информационно-измерительная система, которая работает по спектрам объектов.

12. Планы на следующее занятие

И есть ещё возможность, э-э другие ещё есть возможности использования графических интерфейсов. Как я вам говорил, что есть э-э спектральный анализ, да? А есть ещё я не показал, где находятся э-э инструкции вот эта по разработке приложений собственных. Вы эту инструкцию открываете, начинаете искать исходные данные, устанавливаете систему и начинаете пробовать решать задачи. Вот. Показываете мне, как у вас получается. То есть на следующем занятии мы начинаем проводить лабораторные работы в таком режиме. Я не показываю вам и рассказываю что-то, а э-э вы мне показываете и рассказываете, что вы делали, а я вам даю советы. И ещё второй вариант есть, что мы э-э всё-таки будем рассматривать работу 3.04. Это реальная работа. То есть она, вот те, что мы до этого проходили, это такого учебного характера. А 3.04 - это реальная работа, за которую даже, за такие работы даже деньги можно заработать, получить, разработка реальной методики риэлторской оценки недвижимости или там машин с пробегом. То есть это реальная задача, не учебная. Но её можно использовать в качестве учебной. Так что вот вы выбираете. Наверное, всё-таки, наверное, э-э 3.04 работу рассмотрим, а потом начнём изучать, э-э создавать свои приложения, ребята. Всё, конец занятия. Надеюсь, вам было интересно. А если нет, ну тогда потом пи-пи-пи-пи-пип, знаете, да, вот это? Вот, всё, до свидания, ребята.

До свидания. До свидания.