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

***Российская Федерация***

**187 Лекция №11. По дисциплине Интеллектуальные информационные системы и технологии. 2020-11-07**

Лекция 11: Основы Систем Распознавания Образов (СПО) и перспективы ИИСиТ (самостоятельно)

Резюме

Введение:

Лекция №11 по дисциплине "Интеллектуальные информационные системы и технологии" (ИИСиТ) прочитана профессором Луценко Е.В. 7 ноября 2020 года. Отмечена низкая явка студентов (около 10%), лектор призывает активнее посещать занятия.

План лекции:

Перспективы развития ИИСиТ: Тема вынесена на самостоятельное изучение. Студентам предоставлена ссылка на соответствующие материалы на сайте лектора (раздел "Кредо").

Системы распознавания образов (СПО): Основная тема лекции, являющаяся частью главы 4 "Описание интеллектуальных систем". Рассматривается учебный вопрос 4.2.

Системы распознавания образов (СПО) - Учебный вопрос 4.2:

Структура вопроса: Включает 6 подвопросов: основные понятия, проблема распознавания, классификация методов, применение (идентификация, прогнозирование), сравнение идентификации и прогнозирования, роль в управлении сложными системами, методы кластерного анализа.

Основные понятия и определения СПО (4.2.1):

Определение и функции: СПО – класс систем ИИ, формирующих конкретные образы объектов и обобщенные образы классов. Основные функции: обучение (с учителем/без), распознавание (идентификация/прогнозирование), оценка адекватности модели (верификация).

Признаки, образы и фазовое пространство: Объекты описываются признаками (измеренными свойствами) на различных шкалах (номинальные, порядковые, числовые/отношений). Признаки формируют многомерное абстрактное "фазовое пространство", которое, в отличие от физического, обычно не является ортонормированным и метрическим. Объект представляется точкой или объемом в этом пространстве.

Обучающая выборка: Подмножество генеральной совокупности, используемое для обучения модели. Должна быть репрезентативной (адекватно отражать всю совокупность) и эргодичной (закономерности должны сохраняться во времени). Возможно "взвешивание" данных для приближения распределения выборки к генеральной совокупности.

Основные операции:

Обобщение (Индукция): Формирование обобщенных образов классов из конкретных примеров. Позволяет выявить значимость признаков.

Распознавание (Дедукция): Сравнение объекта с другими объектами или обобщенными образами с использованием меры сходства (интегрального критерия) для идентификации или прогнозирования.

Мера сходства (Интегральный критерий): Ключевой элемент распознавания. Может быть метрической (основанной на расстоянии, например, Евклидовом) или неметрической (например, корреляция, свертка). Неметрические меры часто предпочтительнее из-за природы фазового пространства и устойчивости к шуму. Система Эйдос использует неметрические критерии.

Самообучение (Кластерный анализ): Автоматическое формирование кластеров (групп) сходных объектов из неклассифицированной выборки.

Комбинированное обучение: Предложен метод, сочетающий обучение с учителем и без (когнитивная дивизивная кластеризация), создающий подклассы для повышения детализации и точности модели.

Верификация, Адаптация и Синтез модели:

Верификация: Оценка адекватности (достоверности, валидности) модели. Необходима перед использованием. Может быть внутренней (на обучающей выборке) и внешней (на контрольной), интегральной (общей) и дифференциальной (по классам).

Адаптация: Количественное улучшение модели путем добавления новых данных из той же генеральной совокупности.

Синтез (Пересинтез): Качественное изменение модели при появлении новых закономерностей, объектов или классов, требующее введения новых признаков или шкал.

Детальная расшифровка текста

Введение

Здравствуйте, ребята. Не здороваетесь. Надо здороваться. Тем более запись идёт. Здравствуйте, ребята. Ёлки-палки. Запись идёт же не чата, а микрофон. Ладно. Проехали.

Сегодня 7 ноября 2020 года. Это суббота, что интересно. И первая пара. 8:00-9:30. Лекция 11 по дисциплине Интеллектуальные информационные системы и технологии. Занятие ведёт профессор Луценко Евгений Вениаминович.

Низкая явка и призыв к активности

Ребят. Значит, теперь давайте, пожалуйста, как-то, я не знаю, будите своих товарищей, потому что просто нехорошо, что вас нету.

У вас в одной группе 25 человек, в другой группе 21 человек, и в третьей группе тоже 25. То есть вас здесь должно быть 71 человек. Понимаете? А вас, знаете сколько? Семь. То есть вместо 71 - семь. Это 10%. Это полный отстой. Давайте, звоните, будите там всех. Пусть проявляются. Лекция.

Причём и старост нету даже, я так смотрю, да?

Ну-ну, давайте, давайте, как-то в группу там пишите, я не знаю. Что такое? Семь человек. Сейчас восемь стало. Из 71.

Я понимаю, что суббота, утро. Понимаю, что вчера была пятница, вечер. Ну и что? Потом будет четверг, день. Ну и что теперь из этого? Всё время что-нибудь не то: то дождь, то нет дождя, да? То утро, то вечер. Всё время это вам причиной является, почему вы не приходите.

О, Господи. Кошмар какой-то.

План лекции

Ребята, значит, у нас сегодня два вопроса.

Перспективы развития ИИСиТ (самостоятельное изучение)

Один вопрос - это перспективы развития информационных интеллектуальных технологий. Что касается вот этого первого вопроса о перспективах, то я оставляю вам его на самостоятельное изучение. Сначала вам дам ссылку на то место, где об этом говорится.

Вот этот материал, который я сейчас дал ссылочку. Большая просьба передать всем, кто отсутствует. Прямо возьмите, передайте вот сейчас, перешлите в группу, там, куда хотите. Перешлите. Значит, как его найти этот материал? Смотрите. Значит, выходим на мой сайт. Мой сайт - это понятно, да? То есть вы можете на телефончике выйти на мой сайт, войти по второй пункт. И здесь, здесь, вот здесь вот, где картинка, тоже опять же, тут все все эти вот находятся ссылки основные. Тут есть кредо, кредо. Вот это кредо нажимаете и посылаете.

Так что вот так вот. Это самостоятельная работа. Но я так надеюсь, вам будет интересно это. Причём я действительно так думаю, что это будет интересно. Не потому, что вот я занятие веду, а просто действительно это интересный материал. Но я не буду его освещать, потому что вас очень мало. Может, когда-нибудь, когда вас будет побольше, расскажу про это. Особенно, если у вас вопросы возникнут. Материал очень интересный, касающийся перспектив.

Системы распознавания образов (СПО)

И второй вопрос - это мы начинаем рассматривать конкретные виды интеллектуальных систем. И сегодня мы рассматриваем систему интеллектуальную, то есть вид, вид систем, системы распознавания образов.

Системы распознавания образов (СПО) - Учебный вопрос 4.2

Теперь я открыл вам экран, и вы видите тему сегодняшнюю. Вопрос учебный 4.2. Автоматизированные системы распознавания образов. Мы будем и дальше двигаться. Пройдём этот вопрос, начнём следующий рассматривать и так далее.

Этот вопрос является частью, ребята, темы большой, которая называется виды, виды систем распознавания образов. Описание... Нет, извините. Тоже, видно, не до конца ещё проснулся. Это раздел главы 4. Описание интеллектуальных систем. И мы первый раздел уже рассмотрели этой главы, 4.1. Это системы с интеллектуальной обратной связью и интеллектуальным интерфейсом. То есть там довольно много было интересных моментов, наверное, для вас. И вот теперь мы продолжаем рассматривать вопросы учебные этой главы.

Да, ребята, я вам сказал, что занятие ведёт профессор Луценко Евгений Вениаминович, нет? Да. Занятие ведёт профессор Луценко Евгений Вениаминович.

Следующие вопросы, которые будут у нас, это автоматизированные системы распознавания образов, вот сегодня. Потом математические методы и автоматизированные системы поддержки принятия решений. Потом автоматизированные системы...

Следующий вопрос будет экспертные системы, нейронные сети, генетические алгоритмы, моделирование биологической эволюции, системы когнитивного моделирования, выявление знаний из опыта, интеллектуальный анализ данных. И области применения интеллектуальных технологий, перспективы их развития, в том числе в интернет. Это интереснейшие вопросы, и очень важные. Мы будем их рассматривать вот в последующих лекционных занятиях.

И потом будет у нас глава 5, следующая. Это модели представления знаний. Это ключевой вопрос по вашей дисциплине, который вот необходимо обязательно вам иметь представление об этом.

Ну и начинаем мы изучение вопроса учебного 4.2. Автоматизированные системы распознавания образов.

Структура вопроса 4.2

Здесь у нас есть шесть учебных вопросов:

Основные понятия и определения, связанные с системами распознавания образов;

Проблема распознавания образов;

Классификация методов распознавания образов;

Применение распознавания образов для идентификации и прогнозирования;

Сходство и различия в содержании понятий идентификации и прогнозирования;

Роль и место распознавания образов в автоматизации управления сложными системами;

Методы кластерного анализа.

Хотя вот этот вопрос методы кластерного анализа, он, похоже, что сюда попал несколько неоправданно. И он относится скорей к интеллектуальному анализу данных.

Основные понятия и определения СПО (4.2.1)

Определение и основные функции СПО

Основные понятия. Системой распознавания образов называется класс систем искусственного интеллекта, которые обеспечивают:

Формирование конкретных образов объектов и обобщённых образов классов. Но не все модели обеспечивают формирование обобщённых образов классов.

Обучение с учителем и самообучение

Обучение, то есть формирование обобщённых образов классов на основе ряда примеров объектов, классифицированных, то есть отнесённых к тем или иным обобщающим категориям учителем и составляющих обучающую выборку. Кластерный анализ тоже сюда относится. Обучающиеся системы.

Самообучение, то есть формирование кластеров объектов на основе анализа неклассифицированной обучающей выборки.

Распознавание и оценка адекватности

Распознавание, то есть это идентификация или прогнозирование состояния объектов, описанных признаками, друг с другом или с обобщёнными образами классов.

Определение степени адекватности модели.

Решение обратной задачи идентификации и прогнозирования обеспечивается не всеми моделями распознавания образов.

Вот это основные функции системы распознавания образов. И в это всё вопрос основные понятия 4.2.1.

Признаки, образы и фазовое пространство (4.2.1.1)

Он довольно большой.

Определение признаков и свойств

Что такое признаки и образы конкретных объектов? И метафора фазового пространства. Признаками объекта будем называть конкретные результаты измерения значения их свойств. То есть у объектов и их состояний есть свойства, и эти свойства фиксируются, измеряются, фактах наличия этих свойств. И вот этот, это наличие, и есть, собственно говоря, признак. Ну, например, свойство - цвет, а признак - красный. Свойство или, допустим, свойство - вес, а значение этого свойства - 60 кг, к примеру. Или свойство - цвет, а значение свойства - красный.

Измерение свойств и шкалы

Свойства объектов отличаются своим качеством и измеряются с помощью различных органов восприятия или измерительных приборов в различных единицах измерения, различных типах измерительных шкал и в различных единицах измерения.

Сразу же вам дам ссылочку сейчас на... Потихонечку просыпаются, да? Дам ссылочку на одну свою статью понятийную такую, основополагающую, которая называется метризация шкал. Это очень важная статья. Если её прочитаете, вы многое поймёте из этой области.

Результатом измерения или наблюдения является снижение неопределённости в наших знаниях о значении свойств объекта. Значение свойств конкретизируется путём их сопоставления определённым градациям соответствующих измерительных шкал: номинальных, порядковых и числовых.

В номинальных шкалах отсутствует отношение порядка, начало отсчёта, единица измерения. Единственное, что есть в номинальных шкалах - это отношение эквивалентности. На порядковых шкалах определены отношения больше-меньше, но также отсутствует начало отсчёта, единица измерения. На шкалах отношений определены и отношения порядка, и все арифметические операции, есть начало отсчёта и единица измерения.

Метафора фазового пространства

Можно представить себе, что шкалы образуют оси координат некоторого абстрактного многомерного пространства, которое будем называть фазовым пространством. Это традиционное название, в физике так называют пространство, где не только там пространственные, временные координаты, но и какие-то другие величины, там температура и давление, например, могут там быть и другие. Какие-то физические величины друг с другом как-то связаны.

Представление объектов в фазовом пространстве (точка или объем)

Можно представить себе... Значит, в этом фазовом пространстве каждый конкретный объект представляется определённой точкой, имеющей координаты, соответствующие значениям его свойств по осям координат, то есть градациям описательных шкал. На самом деле точкой или объёмом. Почему не точкой, а объёмом, ребят? Потому что по каждой оси есть определённая точность измерения, погрешность какая-то. То есть фактически координатами этого объекта являются не точечные значения на осях вот на этих шкалах, а интервальные значения. То есть это точка плюс какая-то конкретная, и плюс погрешность измерения, например. Или какой-то числовой диапазон. А это даже в числовых шкалах. А в порядковых там вообще само собой понятно, что там интервальные значения текстовые. Там никак не может быть точечного значения.

Свойства фазового пространства (неортонормированность, неметричность)

Вы все привыкли к тому, что пространство является ортонормированным, потому что ортонормированным является наше естественное физическое пространство. То есть наши, в нашем пространстве, в котором мы ориентируемся физическом, все оси координат взаимно перпендикулярны, если декартову систему координат использовать. Следовательно, по значениям на одной оси невозможно посчитать значение по другой оси, потому что они никак друг с другом не связаны. Если бы там между ними был угол не равный 90°, а какой-то другой, то можно было бы просчитать там, используя синус, косинус там эти углы. А так, то есть эти значения на других осях. А так это сделать невозможно. То есть это независимые друг от друга оси в обычном пространстве. А вот в этом абстрактном фазовом пространстве, в общем случае, оси не являются взаимно перпендикулярными. То есть, в общем случае, это пространство не ортонормировано. Более того, оно и не метрическое.

Обобщенные образы классов (4.2.1.2)

Признаки обобщённые образы классов. Обобщённый образ класса формируется из нескольких образов конкретных объектов, относящихся к данному классу, то есть одной градации некоторой классификационной шкалы. Обобщённые образы классов формализуются, кодируются путём использования классификационных шкал и градаций, которые могут быть тех же типов, что и описательные, то есть номинальные, порядковые и отношений и числовые. Видите, я всё время говорю о шкалах, поэтому я вам сослался на эту вот статью о шкалах.

Обучающая выборка: Репрезентативность и Эргодичность (4.2.1.3)

Определение и значение обучающей выборки

Понятие обучающей выборки, её репрезентативность по отношению к генеральной совокупности, ремонт и взвешивание данных. Это вопрос 4.2.1.3. Обучающей выборкой является некоторое подмножество исследуемой совокупности, которая называется генеральной совокупностью. На основе изучения обучающей выборки мы хотели бы сделать выводы о генеральной совокупности. Причём важно знать степень достоверности этих выводов.

Репрезентативность

Рассмотрим, как зависит степень достоверности выводов о генеральной совокупности от объёма обучающей выборки. Если обучающая выборка включает все объекты генеральной совокупности, то есть они совпадают, то достоверность выводов будет наиболее высокой при всех прочих равных условиях. Если же обучающая выборка очень мала, то вряд ли на её основе могут быть сделаны какие-то достоверные выводы о генеральной совокупности, так как в этом случае в обучающую выборку могут даже не входить примеры объектов всех или подавляющего большинства классов. То есть она будет слишком фрагментированной. Под репрезентативностью обучающей выборки понимается её способность адекватно представлять генеральную совокупность, так что изучение самой генеральной совокупности можно корректно заменить исследованием обучающей выборки. Репрезентативность - это очень важное понятие, которое я вам сейчас сформулировал.

Эргодичность

Следующий момент - понятие эргодичности. Я ввожу понятие эргодичности. Ещё здесь оно отсутствовало. Значит, аналогично понятию репрезентативности, но для времени. Что я могу сказать? Что если мы исследуем выборку, относящуюся к одному времени, то насколько корректно на основе неё делать выводы о другом времени? Относящиеся к определённому периоду времени модели делаем выводы о других периодах времени, то это корректно, если в ходе жизни этих периодов сохраняется эргодичность процессов, то есть действуют те же закономерности. В предметной области действуют те же закономерности. Нарушение эргодичности происходит в точках бифуркации. Это очень важные понятия, ребята, поэтому я их сейчас и нашлёпал, как вы понимаете. Потому что это очень важно. Это то же самое, только во времени.

Ремонт и взвешивание данных

Репрезентативность зависит не только от объёма, но и от структуры обучающей выборки, то есть от того, насколько полно представлены все категории объектов генеральной совокупности, классы, и от того, насколько полно они описаны признаками. Что такое взвешивание данных или ремонт обучающей выборки? Это такая операция, в результате которой частотное распределение объектов по классам в обучающей выборке максимально, насколько это возможно, приближается к частотному распределению в генеральной совокупности, если оно известно или из независимых источников, либо к равномерному. В системе Эйдос режим взвешивания данных реализован. Ещё до с версии. А в этой версии под Windows ещё он не реализован.

Основные операции: Обобщение и Распознавание (4.2.1.4)

Теперь следующий вопрос 4.2.1.4. Основные операции, которые выполняются системой распознавания образов. Это обобщение и распознавание.

Обобщение (Индукция)

Обобщение, индукция - это операция формирования обобщённых образов классов на основе описания конкретных объектов, относящихся к этим классам, входящих в обучающую выборку. Сразу необходимо отметить, что операция обобщения реализуется далеко не во всех моделях распознавания образов. Например, в методе К-ближайших соседей. Например, её нет в методе К-ближайших соседей. А в тех, в которых она реализуется, это делается по-разному. Обычно, пока не реализовано обобщение, нет возможности определить ценность признаков для решения задачи идентификации. Например, если у нас есть 10 конкретных мячей разного размера и цвета, состоящих из разных материалов и предназначенных для разных игр, а мы рассматриваем их как совершенно независимые друг от друга объекты, наряду с другими, то у нас нет возможности определить, какие признаки являются наиболее характерными для мячей и наиболее сильно отличают их от других объектов. Но как только мы формируем обобщённый образ мяч, стул и другие, сразу выясняется, что цвет мяча и материал, из которого он сделан, не являются жёстко связанными с обобщённым образом класса мяч. А наиболее существенно то, что он круглый, и его можно бросать или бить во время игры. То есть это степень характерности признаков определяется именно после обобщения. Если обобщения нет, то и тогда и не определишь степень характерности признаков для того или иного объектов той или иной категории, того или иного класса.

Распознавание (Дедукция) и Мера Сходства (4.2.1.5)

Определение распознавания

Что же такое распознавание, ребята? Это операция сравнения и определения степени сходства образа данного конкретного объекта с образами других конкретных объектов или с обобщёнными образами классов, в результате которой формируется рейтинг объектов или классов по убыванию сходства с распознаваемым объектом. Вот если, допустим, берём студентов ваших, да, вашей группы, и берём, значит, распознаём этих студентов, то как это будет выглядеть? Можем сказать так, что вот такой-то студент больше всего похож на таких и таких студентов. То есть конкретно можно прямо сказать. И ещё можно сказать, что он похож вообще на обобщённый образ студентов. То же самое касается студенток. И можно взять любого конкретного студента и сравнив его со остальными студентами и студентками этой группы, например, других групп, составить их рейтинг по степени сходства. То есть показать, что этот вот больше всего похож на этого, этот на этого. То есть этот конкретный больше всего похож вот на этих вот, и составить их список в порядке убывания степени сходства.

Наверное, то, что я вам рассказываю, вам смутно напоминает... Что, ребята? Что вам это напоминает, то, что я сейчас рассказываю? Вот анализ. Это напоминает, в общем-то, систему Эйдос. То есть система Эйдос - это многофункциональная интеллектуальная система, имеющая сходство, в частности, и с системами распознавания образов. То есть то, что понимается под системами распознавания образов, всё это она делает. Кроме этого, она делает ещё много другого, то, что связано с когнитивным моделированием, принятием решений. Понятно, да? Анализом данных, интеллектуальным анализом данных. То есть выявлением знаний из опыта и использованием этих знаний для познания, моделирования предметной области. То есть это система такая многофункциональная. Но в частности, вот то, что я сейчас рассказываю про системы распознавания, я же рассказываю не про систему Эйдос, а вообще про системы распознавания. То есть можно взять книжки по этому поводу, почитать и обнаружить, что я вот рассказываю примерно то же самое, что там написано. То есть обычные вещи я рассказываю, касающиеся систем распознавания образов.

Интегральный критерий (Мера сходства)

Ключевым моментом при реализации операции распознавания в математической модели является выбор вида интегрального критерия или меры сходства, который бы на основе знания о признаках конкретного объекта позволил бы количественно определить степень сходства этого конкретного объекта с другими объектами или обобщёнными образами классов.

Метрические и неметрические меры сходства

Очень важно, на мой взгляд, вот, ребята, я сейчас вам рассказываю, конечно, то, что общеизвестно, но в какой-то степени есть авторский вклад в том, что я даю оценки, акценты расставляю какие-то, ваше внимание на какие-то моменты обращаю. И очень многие моменты, они творчески переработаны мной, ну, просто они как бы, скажем, это попроще, ну, определённое понимание у меня сложилось за длительное время работы в этой области. И это понимание я пытаюсь вам донести. То есть не просто формально излагаю некий текст, а просто вот ещё излагаю понимание своё этого. Так вот, я могу вам сказать, что вот эти меры сходства - это ключевой момент. И здесь нужно понимать, что есть метрические меры сходства и неметрические. Метрические меры сходства, они связаны с тем, что в пространстве какая-то есть метрика, способ измерения расстояния. Обычно это теорема Пифагора в евклидовом пространстве. И она вот так вот и применяется, как будто бы вот это пространство, которое создаётся в моделях этих распознавания образов, как будто оно тоже является евклидовым. С какого перепугу, ребята? Вот так вот подумайте. Этих пространств существует миллион разных видов, понимаете? В частности, они могут быть искривлёнными, например, а могут быть вообще топологически не сходными с евклидовым пространством. Ну, скажем, пространство на тороидальной поверхности, тор, например. Понимаете? Это что такое? Там, и мы применяем там меры, которые мы привыкли на плоском пространстве применять. Это же полная глупость вообще-то, понимаете?

Устойчивость критерия к шуму

Поэтому очень большую роль играют неметрические меры сходства. Требование устойчивости к наличию шума математически означает, что результат применения интегрального критерия к сигналу, состоящему только из белого шума, должен стремиться к нулю при неограниченном увеличении объёма данных. Это определение белого шума.

Корреляция как мера сходства

Это значит, что в качестве интегрального критерия может быть применена функция, используемая при самом определении понятия белый шум, то есть свёртка или скалярное произведение, корреляция. Корреляция Пирсона - это скалярное произведение стандартизированных векторов в координатной форме. Вот я вам сейчас такие вещи сообщаю, которые вообще мало кто и знает, наверное. Кто знает, что корреляция Пирсона - это произведение стандартизированных векторов в координатной форме? И что, по сути дела, это скалярное произведение в математической своей форме? Я так думаю, что и не очень-то многие знают. А это было бы неплохо знать. Такой интегральный критерий и предложен в математической модели системно-когнитивного анализа и реализован в системе Эйдос.

Самообучение (Кластерный анализ) (4.2.1.6)

Обучение без учителя или ещё его называют самообучение. Это процесс формирования обобщённых образов классов на основе обучающей выборки, содержащей характеристики конкретных объектов, причём только в описательных шкалах и градациях. Этот процесс реализуется в три этапа:

Кластерный анализ объектов обучающей выборки, в результате которого определяются группы наиболее сходных из них по их признакам и формируются кластеры.

Этим кластерам присваивается статус обобщённых классов.

Для формирования обобщённых образов классов, которые используются конкретные объекты, входящие именно в эти кластеры. И потом формирование обобщённых образов классов, аналогично тому, как это делалось при обучении с учителем.

Комбинированное обучение (с учителем и без)

Предлагается высокоэффективный алгоритм обучения с учителем, совмещённый с алгоритмом когнитивной дивизивной кластеризации, то есть обучением без учителя. Разбивать класс на подклассы с теми же наименованиями, если объект не очень похож на класс. При сравнении с классом сравнивается со всеми подклассами. Использовать объект для обучения наиболее похожего подкласса. Идентификация объекта с любым подклассом означает его идентификацию с классом. При объединении подходов с учителем и без учителя, идентификацию объекта с подклассом считать идентификацией с классом. Тут немножко написано по-другому, так как я тогда вот, когда писал это, думал. Но я могу сейчас по-другому ещё сказать. Это при дивизивной кластеризации, то есть разделении классов на подклассы. А вот здесь дальше написано то, что я сейчас хотел сказать. Перед тем, как использовать объект для формирования обобщённого образа класса, определять степень его сходства с этим классом. А, значит, учитель сказал, что этот объект относится к этому классу. Хорошо. Мы берём, значит, не сразу сходу формируем, используем его для формирования образа этого класса, а сначала определяем степень сходства с этим классом, с образом этого класса. Если уровень сходства выше некоторого порога, то использовать этот конкретный объект для формирования этого образа. А иначе, а иначе создавать новый класс с тем же именем и использовать для его... этот объект для его формирования. Использовать образ объекта для его формирования.

Ну, я сейчас вам просто приведу пример, и вы, наверное, тогда более наглядно поймёте. Вот представьте себе, что мы нам нужно сделать систему OCR, да, которая распознаёт тексты. Вот мы берём, сканируем какую-то страницу на сканере, вот, допустим, вот эту. И берём эту страницу в виде вордовского файла, размещаем в специальной папочке. И система OCR начинает учить модель свою, начинает создавать модель, в которой должно быть отражено, какие пиксели...

Допустим, мы нашли буковку А в тексте, вот здесь, которая первый раз встретилась эта буковка. И эксперт нам, то есть не эксперт, а вот в этом тексте у нас там есть просто этот текст. Система распознавания текста определила, где там у нас строчки, где символы в строчках. И подошла к символу А, вот, к первому, который там встретился, вот буква К, например, да, в слове. И видит, что образа этой буквы ещё нет в модели. Вообще её нет такого образа, образа этого класса. Она создаёт этот образ и формирует его на основе данной конкретной реализации этого образа в виде буквы А, вот этой напечатанной вот здесь. Потом, когда опять встречается буква А, вот, допустим, вот здесь, она встретилась в слове формирования, то она не сразу её использует для для продолжения формирования образа буквы А, а сначала сравнивает, а похожа эта буква А вторая? Она тоже имеет тот же самый код ASCII или там TrueType Fonts, да, этот код тот же самый. Значит, она смотрит, использовать этот образ для формирования этой буквы А или нет. Значит, если эта буква А похожа на первую выше некоторого порога, ну, допустим, больше 90%, тогда используется этот образ, конкретно этой буквы А для формирования обобщённого образа буквы А, который первый раз начал был создан и начал формироваться. А если она не похожа? Вот, допустим, у нас вот здесь слово верификация, и здесь тоже буква А. Но она не похожа вот на ту первую. Почему она не похожа? А потому что она, видите, полужирным шрифтом сделана. А вот здесь вот тоже у нас есть буква А, только она курсивом, не полужирным, но курсивом. А вот здесь ещё к тому же она и жирная, и полужирная, и ещё курсивом. Что это значит? Это значит, что если мы начнём формировать образ буквы А, вот этот, который был сначала создан, на основе этих всех примеров, и полужирного, и курсива, и жирно-полужирного курсива, то у нас будет искажённый образ буквы А. То есть она будет плохо идентифицироваться. Почему? Потому что вообще-то объекты, которые были предъявлены для формирования этого обобщённого образа, довольно заметно отличаются друг от друга. Поэтому, если мы обнаружили вот эту полужирную букву А и производим идентификацию с этим обобщёнными образами, понимаем, что это не очень похоже на тот образ буквы А, который до этого у нас формировался. Мы тогда берём и создаём новый образ буквы А, ещё один. И пишем там образ буквы А, там, ну, с, допустим, -1, -2, -3. Ну если первый раз, допустим, это был просто образ буквы А, теперь мы берём -1 пишем. И формируем его. Потом берём ещё просто курсивом. Тоже А-3 там, например. Полужирный курсив, значит, А-4 там и так далее. У нас возникнут десятки там разных видов классов, обозначающих все разные буквы А, но написанные разными шрифтами, разных размеров, с разными стилями, полужирные, там курсивы, там сочетания этих стилей и так далее, подчёркнутые, underline там и так далее и тому подобное, и так далее. В результате, когда система распознавания будет распознавать текст, она не просто будет распознавать, что это буква А, ребята, она ещё будет распознавать, каким шрифтом она написана и в каком стиле. Улавливаете, ребят? Скажите, улавливаем или там не улавливаем? Скажите. Понимаете, о чём я говорю вообще? Да, понимаем, улавливаем. О, отлично. Ну здесь пришли самые лучшие, поэтому они и понимают. А те, кто спит, спокойного сна, да им.

Верификация, Адаптация и Синтез модели (4.2.1.7)

Верификация (Оценка адекватности)

Что такое верификация, адаптация и синтез модели? Я, значит, думаю, что, значит, что как только произносится слово модель, сразу же должно должен возникать вопрос о степени её адекватности. Значит, я хотел бы, ребята, чтобы вы это хорошенько запомнили. Вообще просто вот взяли вот, ну как вот, я не знаю, золотыми буквами такими огромными вписать надо в ваше подсознание и в сознание, и в подсознание. Что как только кто-то где-то заговорил о модели, на какой-нибудь конференции, в какой-нибудь статье, книжке, на защите какой-нибудь там, сразу же вы должны вспоминать, что я вам говорил. Как только кто-то сказал слово модель, вы сразу же должны слово адекватность вспоминать. Модель - адекватность. Они должны связаны быть, понимаете, друг с другом. То есть, если модель есть, то можно ли ей пользоваться? Вот в чём вопрос. А может быть, и нельзя ей пользоваться. Если она недостоверна, то ей пользоваться нельзя. Понимаете? Почему? А потому что неверные будут результаты прогнозирования, неэффективные будут приниматься решения, даже, может быть, опасные какие-то решения будут приниматься, которые вместо того, чтобы улучшить ситуацию, могут её ухудшить. Потому что модель-то неадекватная, она же неправильно отражает предметную область. Поэтому этой моделью пользоваться вообще нельзя. А вот измерение адекватности - это и есть верификация. Что такое верификация? Верификация модели - это операция установления степени её адекватности или валидности, говорят, путём сравнения результатов идентификации конкретных объектов с их фактической принадлежностью к обобщённым образам классов. Для цели верификации модели могут быть использованы и задачи прогнозирования и принятия решений, но это более проблематично. В чём заключается проблематичность, ребят? В том, что при идентификации у нас все данные есть в обучающей выборке, налицо, как говорится. Вот они у нас есть, мы их просто используем и всё. А при прогнозировании мы прогнозируем, да, что-то, а потом ждём, когда это осуществится, понимаете? То есть это уже сложнее оценить достоверность модели. Можно, конечно, использовать ретроспективные данные, то есть прогнозировать то, что уже произошло. Но тогда это мало чем отличается от идентификации. То же самое касается принятия решений. Но принятие решений ещё более сложно. Если мы принимаем решение, то нужно потом это решение ещё и выполнить, реализовать его, а потом смотреть, получился результат, который мы хотели или нет. И вот если он получился, значит, модель достоверна, если не получился, значит, недостоверна. Ну это примерно, как вот оценивать, допустим, правильность теории Маркса. Берём мы, совершаем революцию. Сегодня как раз у нас годовщина революции, правильно? Вот я вспомнил про революцию. Это день Великой Октябрьской социалистической революции, которая произошла в 1917 году. По новому стилю это 7 ноября она произошла. Вот, значит, мы можем посчитать: 2021 - 1917 = 104 года назад произошла эта революция. Её по-разному называют: революция, переворот. Ну, в любом случае, можно сказать так, это было качественное изменение общественно-политического строя, это точно. И могу вам сказать, что она произошла не от нечего делать, не от каких-то там фантазий, а это было глубоко обусловленным процессом, таким глубинным. И все остальные процессы исторические, которые потом происходили: и Великая депрессия, и Первая мировая война, и Вторая мировая война, они все связаны с этими глубинными процессами. И революция, это Октябрьская революция - это был один из ответов, которые выработало общество на эту вот реальность, на развитие реальности. То есть это было обусловленным процессом. То есть до него общество дошло до того, что пришло к этому результату. Потом эти же проблемы, которые пытались решить путём этой революции, они их же люди пытались решить и другими способами, более или менее успешно, в какой-то степени. Я появляюсь, допустим, в 2000 там 16 году и принимаю участие в этих спорах с Плехановым, Лениным. Я высказываю такую мысль. Слушайте дальше внимательно, какую мысль я им высказывал. Они вообще её проигнорировали полностью эту мысль. Я им говорил так: "А вы вообще..." Они спрашивают меня: "А вы как думаете, достоверно учение Маркса или нет?" А я говорю: "Вы знаете, я бы проверил бы сначала. Взял бы, проверил бы, ну, на кроликах там, на мышах как-нибудь там, не знаю. Вот у них бы взял бы, совершил бы революцию в каком-нибудь там миллиметровом государстве, которое на карте не видно. Там в Литве где-нибудь, там, в Эстонии. Не так жалко будет, понимаете, как Российская империя, величайшая, так сказать, империя мира, которая занимает шестую часть Земли. Вот, почти со всеми океанами имеет границы там. И в общем, говорят, региональная держава. Барак Обама сказал: "Региональная держава Россия". Вот да, она региональная, но регион - весь мир. Она граничит с Америкой, она граничит с Европой, она граничит с Азией. Она имеет океаны, которые на Северном полюсе, так сказать, находятся, эти полюс находится, северно-ледовитый океан, имеет южные моря в границах своих, и Атлантический океан, и Тихоокеанский бассейн. Нет другой такой региональной державы, которая весь мир охватывает, понимаете, кроме России. Вот. Так вот, такая она региональная, но регион очень большой, понимаете? Больше того, он не просто очень большой, он самый большой в мире, больше, чем у любой другой страны. Так, между прочим, так сказать. Вот взяли бы, проверили на кроликах. Получилось у них хорошо. Ну тогда, значит, у нас возникает такое предположение, что эта теория адекватна. Давай тогда масштабировать её с учётом масштабов. Может ли это реализоваться в государстве больших масштабов? Ну, допустим, в Польше или там в Чехии, или там в Венгрии. А почему я про них вспоминаю? Потому что у них-то это не очень-то они были за то, чтобы это реализовать. И даже в самой Германии, в этой Восточной Германии, там очень у них хорошо получилось. Вот у них, пожалуй, получилось лучше, чем у всех остальных. Ну эта теория-то немецкая вообще-то, из Германии она ж пришла. Это примерно, я сейчас вспоминаю фразу, когда Ангела Меркель сказала Путину: "Ваш Маркс". Вот ваш Маркс, вот из-за него там столько вот было проблем. А он говорит: "Вообще-то он не наш Маркс, вообще-то он ваш". Это Путин ответил. Мне очень понравилось. Вообще-то он ваш Маркс, понимаете, дорогая Ангела. Вот. Так вот, проверка адекватности модели путём прогнозирования, принятия решений, она очень трудоёмка и опасна. Но применять модель без такой проверки, вообще это просто авантюра, понимаете, авантюризм, профанация всего вообще учения о том, что какие-то модели можно использовать для чего-то там, для принятия каких-то решений. Ну лучше всего, проще всего проверять достоверность модели на способности этой модели решать задачи идентификации. То есть верификация осуществляется путём идентификации, отнесения объектов к тем или иным классам. Это проще всего, ребят. Поэтому так обычно и делается. И именно это и предложил Ван Рисберген, Ван Рисберген в своей знаменитой F-мере Ван Рисбергена.

Различают внутреннюю валидность, внешнюю валидность, интегральную и дифференциальную валидность. Что такое валидность, ребята? Я сказал, верификация модели, установление её адекватности или валидности. В русском языке слово валидность не применяется, но применяются слова, которые от него производными являются. Скажем, инвалид. Кто такой инвалид? Человек, у которого есть какие-то нарушения, да, дефекты физические, физиологические. А что это слово означает в переводе? Оно означает правильность. Просто-напросто. То есть адекватная модель - это правильная модель, модель, которая правильно отражает предметную область, и всё. И вот измеряют эту валидность. Этот термин валидность используется больше всего психологами в области психологии.

Различают внутреннюю валидность, внешнюю, интегральную, дифференциальную. И, значит, это очень интересные понятия. В системе Эйдос все они реализованы эти понятия, которые сейчас я излагаю. Все они там в разных режимах есть их реализация, отражение. Значит, смотрите, ребята, значит, внутренняя валидность - это способность модели верно идентифицировать объекты обучающей выборки. Есть такое в системе Эйдос? Вот скажите микрофоном. Есть такое или нет? Есть. Итак, чаще всего и делается. Значит, если модель имеет низкую внутреннюю валидность, это в режиме 3-4 делается, то модель вообще нельзя считать удачно сформированной. Ну как можно надеяться, что она там что-то отражает, если она даже обучающую выборку не может идентифицировать правильно? Внешняя валидность - это способность модели верно идентифицировать объекты, не входящие в обучающую выборку, но входящие в генеральную совокупность, по отношению к которой обучающая выборка репрезентативна.

Адаптация модели

Адаптация модели - это учёт в модели объектов, не входящих в обучающую выборку, но входящих в генеральную совокупность, по отношению к которой данная обучающая выборка репрезентативна. То есть это количественное улучшение модели. Если моделью верно идентифицируются объекты, не входящие в обучающую выборку, то это означает, что эти объекты входят в генеральную совокупность, по отношению к которой данная обучающая выборка репрезентативна. Следовательно, на основе обучающей выборки удалось выявить закономерности взаимосвязи между признаками и принадлежностью объектов к классам, которые действуют не только в обучающей выборке, но и имеют силу и для генеральной совокупности. Адаптация модели не требует изменения классификационных и описательных шкал и градаций, а лишь связана с изменением обучающей выборки и приводит к количественному изменению модели, количественному изменению модели.

Синтез (Пересинтез) модели

Что такое синтез, ребята, или повторный синтез, пересинтез модели? Это учёт в модели объектов, не входящих ни в обучающую выборку, ни в генеральную совокупность, по отношению к которой данная обучающая выборка репрезентативна. Эти объекты с новыми, ранее неизвестными закономерностями взаимосвязи признаков с принадлежностью этих объектов к тем или иным классам, причём и признаки, и классы могут быть как те, которые уже были отражены в модели ранее, так и новые, требующие для формализации новых классификационных и описательных шкал и градаций. Пересинтез модели приводит к качественному изменению, качественному её изменению.

Заключение

Мы остановились на вопросе 4.2.2, который называется проблема распознавания образов. Я пишу себе на следующем занятии лекционном. Это у нас будет на следующем занятии. Проблема распознавания образов и так далее, последующие вопросы, которые будут у нас по учебному пособию. Учебное пособие, ребята, размещено в ResearchGate, Moodle, на учебной системе нашей дистанционного обучения университетской. Там, правда, в Moodle там ссылочка дана только. Так что мы будем дальше всё это изучать. Но я могу вам сказать, ребята, что объём информации огромный. И если пытаться её всю изложить, то это, скорее всего, невозможно в те временные сроки, тех те сроки, которые у нас отведены нашим учебным планом. Поэтому я буду многие вопросы оставлять на самостоятельное изучение вам, что я сегодня и сделал. Так оно и предусмотрено вообще-то. Или буду как-то в сокращённой форме их излагать.

Всё, ребята, до свидания. Всего самого хорошего. Передайте самое ценное из того, что я вам сказал сейчас, вот эти ссылочки: ссылочку для самостоятельного изучения, ссылочку на статью и вот эту ссылочку на вторую статью про метризацию шкал и которая в ResearchGate. Вот почитайте, пожалуйста, сами и дайте эти ссылки остальным. Ну, можно старосте дать и сказать, чтобы он, так сказать, или она там, кто там она. Ну да, старостам. Это ж лекция, всем старостам. Вот, чтобы они передали учащимся, студентам. Всего самого-самого хорошего. До свидания. До свидания, скажете? До свидания. Ну, вообще классно. До свидания. До свидания.