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

***Российская Федерация***

**173 Теория информации, данные и знания. Лекция 9. Методика вычисления количества информации 2020-10-29**

## Заголовок

Расчет количества информации методом Харкевича в системе Эйдос: Лекция по теории информации, данных и знаний

## Резюме текста (Summary)

Лекция посвящена методике вычисления количества информации о прошлых и будущих событиях с использованием метода Харкевича в рамках системы Эйдос. Лекция является частью курса "Теория информации, данные, знания" и состоялась 29 октября 2020 года.

**Основные концепции и методология (Система Эйдос и АСКА):**
Лектор представляет систему Эйдос как инструмент для преобразования данных в информацию, а информации – в знания. Этот процесс реализуется через Автоматизированный Системно-Когнитивный Анализ (АСКА), включающий этапы:

1. Когнитивно-целевая структуризация и формализация предметной области.
2. Синтез и верификация моделей (статистических и системно-когнитивных).
3. Решение задач прогнозирования и идентификации.
4. Решение задач поддержки принятия решений (обратное прогнозирование).
5. Исследование моделируемой предметной области.

**Формализация предметной области:**
Этот этап включает разработку классификационных (для результатов) и описательных (для факторов) шкал и градаций. Эмпирические данные кодируются с помощью этих шкал для создания обучающей выборки (обычно в формате Excel). Обсуждаются методы обработки числовых (с интервальной дискретизацией) и текстовых данных, а также способы обработки пропущенных значений (нули для чисел, пробелы для текста).

**Синтез и верификация моделей:**
На основе обучающей выборки строятся статистические модели: матрица абсолютных частот (Абс), модели условных и безусловных процентных распределений. Затем синтезируются системно-когнитивные модели, основанные на вычислении количества информации по Харкевичу. Метод Харкевича сравнивает условную (внутри класса) и безусловную (по всей выборке) частоты встречаемости признака, используя логарифм их отношения. Это позволяет оценить информативность каждого признака для классификации. Лектор противопоставляет этот подход (основанный на делении) методу Хи-квадрат (основанному на вычитании). Надежность модели оценивается путем сравнения ее предсказательной силы с моделью, построенной на случайных данных (шуме).

**Решение прикладных задач и исследование предметной области:**
Модели Эйдос применяются для решения задач идентификации, прогнозирования (с использованием аддитивных интегральных критериев), поддержки принятия решений. Исследование предметной области включает использование различных инструментов визуализации и анализа: инвертированные сводные диаграммы, семантический потенциал, кластерный конструктивный анализ классов и значений факторов, нелокальные нейроны и сети, 3D и 2D интегральные когнитивные карты, оценку значимости факторов и степени детерминированности классов. В качестве примера приводится задача прогнозирования землетрясений на основе данных за 120 лет.

**Административная информация:**
Лектор напоминает о предстоящей аттестационной неделе и необходимости сдачи студентами разработанного приложения и его описания в виде научной статьи. Оценка будет базироваться на качестве выполнения этой работы. Подчеркивается важность демонстрации приобретенных знаний, умений и навыков, особенно в условиях дистанционного обучения.

## Детальная расшифровка текста

### Введение

Здравствуйте. Здравствуйте.
Здравствуйте, здравствуйте.
Здравствуйте.

Здравствуйте. Здравствуйте. Проходите пока, да?
Значит, сегодня 29 октября 2020 года.
Шестая пара, 17:20–18:50.

Лекция номер девять.
Методика вычисления количества информации о прошлых событиях, о будущих событиях.
По дисциплине: Теория информации, данные, знания.

И сейчас я вас еще отмечу, ребята, потом уже начнем занятие.

Да, кстати, сейчас только что была лабораторная работа. Если у кого-то есть какие-то вопросы, там вслед этой лабораторной работы, то можете задать. Можете задать.

Так.
Так, сейчас я сделаю демонстрацию экрана.
Буду вам рассказывать.
И, ребята, скажите мне, пожалуйста, вы помните, я вот по своему по своему расписанию смотрю, что у меня вроде как эта тема. Или мы уже ее проходили? Вот мне этот вопрос интересует. То есть вы помните эту тему или нет, чтобы мы ее изучали?

Ну, значит, давайте тогда начнем изучать эту тему. Значит, тема 11.

### 1. Когнитивно-целевая структуризация и формализация предметной области

Методика вычисления количества информации Харкевича на основе перических данных и решение на этой основе задач идентификации, прогнозирования, принятия решений и исследования моделируемой предметной области путем исследования ее модели.

**Учебные вопросы:**

Первый: когнитивно-целевая структуризация предметной области.
Второй вопрос: формализация предметной области, разработка классификационных и описательных шкал и градаций, кодирование с их помощью исходных данных и формирование базы событий обучающей выборки.
Третий вопрос: синтез и верификация моделей. трех моделей статистических: это матрица абсолютных частот, модели условных и безусловных процентных распределений и частных системно-когнитивных моделей на основе частных критериев знаний.
Четвертый вопрос: решение задач идентификации и прогнозирования. Это я вам расскажу здесь про аддитивные интегральные критерии.
Решение пятый вопрос: это решение задач поддержки принятия решений, обратная задача прогнозирования.
Шестой вопрос: решение задачи исследования моделируемой предметной области. И здесь мы рассмотрим, что такое инвертированные свод-диаграммы, что такое семантический потенциал и кластерно-конструктивный анализ классов, кластерно-конструктивный анализ значений факторов, нелокальные нейроны, нелокальные нейронные сети, 3D интегральные когнитивные карты, 2D интегральные когнитивные карты, сравнение классов и сравнение факторов, когнитивные функции, значимость факторов и их значений и степень детерминированности классов и классификационных шкал.
Вот такие у нас вопросы, ребята. То есть вопросов много. Будем их рассматривать сейчас.

Это всё, то, что я сейчас буду рассказывать, делается на примере конкретной задачи.
То есть все эти вопросы я буду рассматривать на примере конкретной задачи.

Ребята, суть того, что происходит в системе Эйдос - это преобразование данных в информацию, а информации в знания.
Данные отражают предметную область, отражают какие-то события предметной области. Мы выявляем, где там описаны в этих данных эти события. Потом выявляем причинно-следственные связи между данными, между событиями, которые описаны в этих данных. И тем самым данные преобразуются в информацию. Потом мы, то есть информация – это осмысленные данные, данные, в которых найдены зависимости. Затем мы э ставим цели э управления, э то есть э какое определяем, какие будущие состояния мы хотим перевести объект управления. И э используя то, что в наших э информационных моделях уже отражены взаимосвязи между прошлыми и будущими событиями, вырабатываем управляющие решения, обеспечивающие достижение цели, то есть управление.
И тем самым информация преобразуется в знания, потому что знание – это информация, полезная для достижения целей.

Какие задачи решаются, в каком порядке при проведении исследования?
Эти задачи являются этапами автоматизированного системно-когнитивного анализа.
Первая задача – это когнитивная структуризация, формализация предметной области. Вторая задача – это синтез и верификация моделей. Третье – решение задач прогнозирования. Четвертое – решение задач принятия решений. И пятое – задача, решение задач исследования моделируемой предметной области.

Первая задача не автоматизирована. Но она и не может быть автоматизирована, потому что суть этой задачи – это понять, что мы будем рассматривать в качестве причин, а что в качестве следствий, что мы будем рассматривать как факторы, а что как результаты влияния этих факторов.
Это когнитивно-целевая структуризация предметной области. Следующий этап – это формализация предметной области. То есть описание в такой форме, которая уже обеспечивает обработку на компьютере.

Что понимается под формализацией предметной области? Разработка классификационных и описательных шкал и градаций и кодирование с их помощью исходных данных. То есть получение обучающей выборки.
Пример того, как выглядит обучающая выборка, ребята. Это чаще всего экселевский файл, но может быть также DBF файл и CSV файл, но чаще всего это экселевский файл, у которого строчки – это наблюдения, а колонки – это э классификационные и описательные шкалы и градации. Они описывают, описательные шкалы описывают, какие факторы действуют на объект моделирования, а классификационные шкалы описывают, какие будущие состояния объект моделирования перешёл под действием этих факторов.
Все эти шкалы могут быть как числовые, так и текстовые. Если шкалы числовые, там должна быть разделитель десятичной части числа должна быть запятая.
Значит, если данные отсутствуют в числовой шкале, то там должны быть нули. При вводе этих данных в систему, есть возможность задать опцию, что нули рассматривать как отсутствие данных. Если данные текстовые, то отсутствие данных – это пробел.

Текст – это более общее понятие, чем число. Число тоже, конечно, является текстом, но не всякий текст является числом, а ещё есть другие тексты. Из этого вытекает такое простое правило или вывод, что если у нас в числовой шкале где-то есть какой-то текст, то все числа будут рассматриваться как тексты уже. Почему? Потому что в базах данных колонка соответствует полю, а поле имеет тип данных. Если, допустим, мы там пишем 298 целых там 7, а потом пишем "привет" слово, то вся эта шкала будет уже как текстовая рассматриваться, и числа будут рассматриваться как тексты уникальные.
Если же э в колонке текстов нет в числовой, то она тогда имеет тип числовой, и рассматривается как числовая. Числовые колонки и текстовые обрабатываются по-разному. Текстовые сортируются в порядке убывания по алфавиту, и при этом э с признаком Unique. Этот признак, он означает, что э если у нас есть несколько раз одни и те же значения в строках в одной колонке, то в результате в индексном массиве оно будет только один раз встречаться.
Числовые колонки обрабатываются по-другому. Они сортируются тоже в порядке возрастания, и берётся минимальное, максимальное значение и делится на число числовых, на количество числовых интервалов или числовых диапазонов, которые задаст пользователь при вводе этой информации. В результате формируется при формализации предметной области формируются описательные шкалы, классификационные шкалы формируются и градации этих шкал.

Вот.

Затем информация вводится в систему Эйдос. В результате получаются классификационные, описательные шкалы и градации и обучающая выборка.
Вот примерно в таком виде это выглядит. Я могу вам сказать, ребята, что все базы данных системы Эйдос открываются в Экселе. Вот этот стандарт данных DBF файл открывается, ну, правда, надо сказать, в старых версиях Экселя открывается. В новых там конвертер убрали оттуда почему-то. Посчитали, что это излишнее.
Значит, в системе есть два варианта выбора диапазонов числовых, два типа числовых диапазонов. Один тип – это диапазоны равные по величине, но с разным числом наблюдений. И второй вариант – это выбор диапазонов, которые разные по величине, но с примерно одинаковым числом наблюдений. Обычно второй вариант, он лучше обеспечивает достоверность модели, чем первый. И вот здесь как раз показано, что у нас немножко отличаются интервалы по величине, иногда они гораздо значительнее отличаются. И примерно одинаковое число наблюдений в них. Ну здесь в данном случае на каждое интервальное значение 947 наблюдений у нас. То есть это довольно большая выборка.
Вот как выглядит обучающая выборка уже в закодированном виде. То есть это те же самые исходные данные, которые были у нас, но уже они закодированы с помощью справочников классификационных и описательных шкал и градаций. И здесь мы видим уже не наименования какие-то и числа, а видим коды соответствующие.

### 2. Синтез и верификация моделей

Дальше осуществляется синтез и проверка на достоверность э системно-когнитивных моделей и э сначала статистических моделей, а потом системно-когнитивных моделей.

Первым делом рассчитывается матрица абсолютных частот, которую ещё называют статистическая модель Abs.
В этой модели колоночки соответствуют классам или обобщающим категориям, а строки соответствуют либо значениям свойств, либо значениям факторов. Что такое значение свойств, значение факторов? Вот представьте себе, что свойство – цвет, а значение – красный. Или фактор, допустим, температура, а значение 25°.
Вот. И система, в системе подсчитывается, сколько раз мы наблюдали факты, э сколько у нас фактов есть наблюдаемых, когда определённое значение фактора действовало на объект моделирования, и он перешёл в определённое состояние. Вот это вот и есть факт. То есть у него было определённое значение свойства, и он относился при этом к определённой категории.

Потом мы на основе этой информации, которая у нас формируется в матрице абсолютных частот, рассчитываем матрицу условных и безусловных процентных распределений.
Это я рассказываю вам о том, как рассчитывается количество информации по Харкевичу на основе перических данных.

Если N итое житое – это у нас число наблюдений итого признака в объектах житой категории, а N житое – это сумма по колонке, то П итое житое – это относительная частота, э условная относительная частота наблюдения итого признака в житой колонке. То есть условная частота встречи этого итого признака у объектов, относящихся к житой, к житому классу. И точно так же мы можем посчитать и безусловную вероятность встречи этого итого признака по всей выборке, у всех объектов выборки. Здесь мы видим в матрице частот, что мы можем посчитать сумму, суммарное число наблюдений этого признаков, всех признаков у объектов житого класса и суммарное число наблюдений итого признака по всем классам. Эти значения потом используются. Ну и суммарное число наблюдений всех признаков по всем классам. Вот. А также можем посчитать, какое количество объектов обучающей выборки было использовано для формирования образа житого класса. Вот. И, соответственно, у нас получается два способа расчёта матриц условных, безусловных процентных распределений, которые здесь как раз у нас и показаны. Первый способ и второй способ. То есть таких матриц две у нас получается.

Здесь мы в таблице 14 видим частные критерии знания, некоторые из которых – это просто частоты, абсолютные частоты, то есть количество встреч итого признака в житом классе, относительные частоты, то есть это э и условные, э и безусловные. Ну, допустим, П итое житое – это частота встречи итого признака у объектов житого класса. И также мы видим здесь частные критерии знаний, которые представляют собой способы сравнения различных способы сравнения фактической частоты наблюдения итого признака в объектах житого класса и теоретической. теоретической частоты наблюдения этого признака у объектов этого класса.

Значит, что касается сравнения, ребята. Значит, слушайте сейчас внимательно, здесь очень интересные моменты есть в этом плане. Значит, я вам сейчас э скажу очень простую вещь, что если у нас есть два числа, то можно их сравнивать либо путём их вычитания друг из друга, либо путём их деления друг на друга. Других способов сравнения никто ещё пока не придумал.
Вот.
Если путём вычитания сравнивать фактическую частоту наблюдения этого признака в объектах житого класса и теоретическую, то это будет э слагаемое Хи-квадрат, элемент матрицы. Если же сравнивать эти э фактическую и теоретическую частоты путём деления, то у нас получится N итое житое умножить на n и разделить на n итое разделить на n житое. То есть получится то выражение, которое находится под логарифмом в мире Харкевича. Если вы посмотрим, что такое П итое житое? Это N итое, N итое житое, разделённое на N житое. Вот оно, видите? А что такое у нас N итое, разделённое на n? А это у нас П итое. Это безусловная вероятность наблюдения этого признака по всем, по всей выборке.

То есть э получается очень интересная штука, что если мы сравниваем фактическую и теоретическую частоты путём вычитания, получается Хи-квадрат. А если путём деления, то получается мера Харкевича. Мера Харкевича – это логарифм от отношения условной и безусловной вероятности встречи этого признака в житой группе и по всей выборке. А чем этот алгоритм интересен? Тем, что если вероятность встречи какого-то признака в группе такая же, как и в среднем по всей выборке, то тогда этот признак никакой информации не содержит о принадлежности или непринадлежности к этому житому классу объекта с этим признаком. Вот. Если признак в группе встречается чаще, чем в среднем, тогда его обнаружение несёт информацию о том, что объект относится к этому классу. Если реже, тогда информацию о том, что он не относится к этому классу.
Таким образом получается, что если у нас никакой информации в признаке нет, то логарифм этот это единица, получается П итое житое равно П итое. Никакой информации в этом признаке не содержится, то логарифм даёт ноль. Если э о принадлежности, то он даёт положительное количество информации. Если о непринадлежности – отрицательное. Это очень здорово. Это соответствует интуитивным представлениям. И означает, что эту меру можно использовать для того, чтобы понять, как действует э итое значение фактора или итое значение свойства на принадлежность э объекта какому-то э классу, обобщающей категории, или переход в какое-то будущее состояние, соответствующее этой категории.

И у нас, когда мы вычитаем, само собой получается, что когда они одинаковые, никакой зависимости нет, э причинно-следственной связи там не выявлены, то тогда получается ноль. А когда отношение, то у нас отношение получается равной единице, если связи нет. И чтобы оно было равно нулю, нужно что-то делать, нормировать каким-то способом. Здесь вот я привожу вам два способа, как это сделать можно. Это взять логарифм, тогда от единицы он равен нулю. А можно просто вычесть единичку из этого отношения. Если вычесть единичку, то мы получаем формулу для известного в экономике коэффициента возврата инвестиций. Если Питое – это инвестиция, а Питое Житое – это возвращённая сумма, то мы видим, что эта сумма возвращённая может быть либо равной инвестициям, либо больше, либо меньше, чем инвестиции. Если она больше, чем инвестиции, то инвестор получает прибыль от своих инвестиций. Если она меньше, то он получает убыток от своих инвестиций. И вот эта величина отношения возвращённой суммы к вложенной минус единица, она, соответственно, имеет знак положительный, если это прибыльное мероприятие было для инвестора, и отрицательный, если это была убыточная предприятие, это инвестирование. При этом я могу вам сказать, что для того, кто получает инвестиции, всегда это выгодно практически. Ясное дело. То есть обычно, кстати, в науке почему-то рассматриваются только интересы инвестора. Никто не рассматривает, куда направлять инвестиции, чтобы получить максимальный эффект в реальной области. Все думают о том, как это сделать, чтобы получить максимальную отдачу, прибыль инвестору. Вся теория управления портфелем инвестиций ориентирована на интересы инвестора. А интересы получателя инвестиций, интересы модернизации, развития реальной области, решения каких-то проблем в реальной области, никого вообще это не волнует, и даже не и даже люди не понимают. Вот когда им говоришь, что вот мы рассматриваем инвестиции как фактор управления, мы рассматриваем инвестиционное управление. Они вообще не понимают, не понимают о чём речь. Мы начинаем объяснять им. Идёт речь о том, что вообще-то инвестиции нужны не для того, чтобы инвестор получил прибыль, как все думают почему-то, а для того, чтобы что-то улучшить в реальной жизни, понимаете? Вот для этого они нужны. Конечно, инвестор при этом не должен страдать. Вот. То есть должен тоже получить прибыль, а не убыток. Или хотя бы свести концы с концами, то есть, ну, выйти на минимальную какую-то прибыль. Ясное дело, что для него это должно тоже быть безубыточным, потому что иначе туда инвестиции не будут направлены.

Вот. Так что вот такие интересные моменты. Оказывается, мера Хи-квадрат тесно связана и с мерой Харкевича количества информации, и с мерой, которая применяется в экономике, возврата инвестиций.
Таким образом, мы сейчас вот увидели, что, э какие у нас есть способы расчёта статистических моделей и моделей э системно-когнитивных, интеллектуальных моделей. Я их называю системно-когнитивные модели.

Значит, вот здесь в формулах есть э коэффициент пси, который я предложил аж в семьдесят девятом году, ребята. И тогда, когда я модель эту разрабатывал, то тогда уже я его посчитал, как он, какое у него выражение аналитическое этого коэффициента. И, значит, я его посчитал из очень простых соображений. Э, чтобы количество информации по Харкевичу, которое мы считаем э при расчёте системно-когнитивной модели, чтобы оно соответствовало в предельном случае детерминистском, в крайнем случае детерминистском, чтобы соответствовало мере мере Хартли. Ну, это требует, наверное, пояснения некоторого, поэтому я сейчас это пояснение вам дам. Значит, вот эта матрица, матрица системно-когнитивной модели. Здесь вот эти коэффициенты в этой матрице, они рассчитываются по формулам, которые приведены вот в этой таблице. В таблице 14, частные критерии. Вот по этой формуле они и рассчитываются. По этим формулам, которые указаны в этой таблице. И получаются разные модели: модели Abs, PRC1, PRC2, INF1, INF3, INF7.

По этим формулам получается, соответственно, 10 моделей: три модели статистические и семь моделей системно-когнитивных.

Сейчас я вам пообещал, что покажу э раскрою э те аспекты, которые связаны с нормировочным коэффициентом линейности и нелинейности.

Да, Николай, привет. Елизавета.
Чат посылаю ссылочку на статью. Статья тринадцатого года. Смотрим на эту статью.

Здесь нарисован график, как действует фактор. Любой фактор на любой объект управления. Если фактор имеет очень маленькое значение, близкое к нулю, то влияние его на объект управления тоже близкое к нулю. При возрастании значения фактора, влияние тоже увеличивается. Вот. Потом, когда значение фактора становится очень большим, то э скажем так, степень увеличения влияния этого фактора уменьшается на какой-то результат, который мы исследуем в объекте управления. Ну, скажем, если мы изучаем, как влияет полив в дозах полива на 1 квадратный метр, на урожайность пшеницы, то мы получаем такого рода зависимость. То же самое касается всех других факторов: удобрений, средств защиты, спашки. Всегда кривая вот имеет вот такой вид, похожий на то, что я сейчас показываю.

Ну, к примеру, значит, возьмём полив. Если полива нет вообще, то урожайность будет э не очень. Если полив увеличивает, то урожайность вроде как увеличивается тоже, растёт. Но потом э при дальнейшем увеличении полива, урожайность уже не растёт. Почему? Как вы думаете? Потому что э слишком хорошо – тоже нехорошо. То есть если, допустим, э эта влага используется растениями, то это нормально. Но если получаются лужи и на поле получаются блюдца воды, то в этих блюдцах пшеница уже не будет, понимаете? А если её залить э пшеницу как рис слоем там, допустим, 20 см, ну тогда всё погибнет, понимаете? Опять будет ноль.

Вот.

То есть фактор обычно влияет нелинейно на объект управления. Теперь смотрим этот коэффициент. Значит, смотрите, у нас если восемь классов, то у нас должно получиться 1 бит информации, если каждому классу соответствует один признак. И этот признак только этому классу соответствует, а другим классам он не соответствует. Это называется детерминистская модель, когда каждому классу соответствует один признак, и больше этот признак никаким классам не соответствует, и этому классу никаких признаков больше не соответствует. Тогда у нас получается три бита информации. Вот чтобы получилось вот здесь вот три, ребята, три получилось чтобы, для этого нужен тот весовой коэффициент, который я вам написал там, показал пси. И если мы возьмём э по два признака на класс, представим себе, то тогда у нас получается по этим вот формулам с этим коэффициентом пси, получается у нас э полтора бита информации.
Вот. То есть смысл такой, что мы больше, чем три бита не можем получить, сколько бы признаков у нас не было. Количество информации максимальное определяется количеством классов и считается по формуле Хартли: логарифм n, где n – число классов.
Вот. Так что вот такие дела. Здесь даже какие-то численные модели в качестве демонстрации приводятся.

То есть я хочу вам сказать, что вот эта модель э расчёта количества информации по Харкевичу, она учитывает нелинейные эффекты.
Если взять, допустим, э один фактор, который увеличивает урожайность на 3%, и другой фактор, который на 3% увеличивает урожайность. То если мы оба эти фактора используем, у нас не получится увеличение урожайности на 6%. То есть результат действия этих факторов не суммируется. Это характерно для нелинейных систем. И даже это является просто определением нелинейной системы. То есть линейная система – это такая система, которая обладает таким свойством, что если на неё действует несколько факторов, совокупность факторов, то результат влияния всех этих факторов вместе, скажем, совместного их влияния, является суммой их влияний каждого по отдельности. То есть один фактор влияет так, другой по-другому. Если они все влияют все вместе, то их результат совместного влияния их равен сумме влияния каждого из них по отдельности. Это линейная система. Так вот этого практически никогда не бывает. Сумма – это аддитивное, значит, влияние, то есть факторы действуют аддитивно на систему, друг с другом не взаимодействуют в этой системе. Это бывает крайне редко. То есть на практике чаще всего получается так, что э результат совместного влияния нескольких факторов не является суммой действия каждого из них по отдельности. А чему равен этот результат? А он равен именно вот по согласно, в соответствии с системно-когнитивной моделью, мы можем определить, чему он равен. Он э то есть мы получаем некое суммарное количество информации о том, что произойдёт с объектом моделирования, когда на него действует совокупность факторов.

И дальше здесь не буду подробно рассказывать. Но могу сказать вам, что сравнение информационной мощности реальной и случайной модели и оценка отношения полезного сигнала к шуму проведена традиционными средствами, теми, которые в теории связи используются. Вот. И у нас получаются хорошие показатели отношения сигнала к шуму. Вот, где-то, ну, больше, чем в два раза. То есть вернее, ну да, где-то восемь раз примерно сигнал превосходит шум. Мощность сигнала превосходит мощность шума. Ну это очень хороший результат, который говорит о том, что в этой сложной области, связанной с землетрясениями, имеется в виду, действительно удалось решить задачи, создать модель, которая позволяет адекватно отразить предметную область и решать задачи.

Ну дальше у нас идёт по плану рассмотрение самих этих задач. Задачи прогнозирования. Вот, это системная идентификация. И здесь мы исходим из чего? Мы уже получили в моделях, в самих моделях у нас отражена информация о том, что произойдёт, если действует то или иное значение фактора. И насколько у нас много информации о том, что это произойдёт. И если у нас действует не один фактор, а много, то предполагается, что произойдёт то, о чём больше всего информации в этой системе факторов. Это называется аддитивный интегральный критерий, когда не один э не один, не одно значение фактора учитывается, а много. Все, какие есть.

В форме, в математической форме это выглядит следующим образом. У нас есть э вектор, описывающий класс, вектор, описывающий состояние объекта э моделирования, которое прогнозируется его будущее состояние. И сравнивается эти два вектора друг с другом. Хорошим способом их сравнения является вычисление скалярного произведения. Обычно, когда речь идёт о двух векторах, обычно все сразу вспоминают, что можно посчитать э корень квадратный из суммы разностей э из суммы квадратов разностей координат этих векторов. Ну то есть что-то вроде теоремы Пифагора, только многомерное обобщение для многомерного пространства, когда много координат у векторов, не три координаты, а много, там десятки, сотни, тысячи, может быть. Так вот, э то есть тысячемерное пространство, а не трёхмерное, короче говоря. Вот. Но дело в том, что вот эта мера эвклидово расстояние, так называемое, которое отражено в теореме Пифагора, она является корректной только для пространства, где э оси этого пространства взаимно перпендикулярны. Или ортонормированы, ещё так говорят, математики называют такое пространство ортонормированное пространство.

Ну, у нас осями пространства являются шкалы, описательные и классификационные. И они между собой взаимодействуют. То есть если посчитать матрицу сходства, то окажется, что они не э ну как, не являются взаимно перпендикулярными. Дело в том, что если мы возьмём, э как считается, какой угол между ними? Вычисляется скалярное произведение, которое как раз вычисляет косинус угла между векторами. Значит, если векторы взаимно перпендикулярны, получается ноль. Если они параллельны, получается единица. Это вот скалярное произведение в векторной форме, это в координатной форме. Значит, если мы возьмём э и стандартизируем векторы, то есть заменим каждую координату её стандартизированным значением, вот, то э у нас получится не скалярное произведение, вот такая формула в координатной форме, а получится вот такая формула. Значит, здесь у нас стандартизировано обычным традиционным путём. вычитание среднего из каждого значения координаты и деление на среднеквадратичное отклонение. Мы видим, что это выражение, оно является выражением для коэффициента корреляции Пирсона.
То есть получается очень интересный вывод, что коэффициент корреляции Пирсона, я не знаю, где-нибудь кто-нибудь вам про это скажет или нет. Вот я вам говорю про это. Что коэффициент корреляции Пирсона, по сути, является скалярным произведением двух векторов, но векторы перед этим предварительно нормируются, стандартизируются. Ну, то есть из них вычитается, из каждой координаты вычитается среднее по координатам вектора и делится на среднеквадратичное отклонение.

Ну, что мы можем сказать? Что если корреляция между классами, э то есть классификационными шкалами или корреляция между описательными шкалами не равна нулю, то это пространство не ортонормированное. И эвклидово расстояние неприменимо. Но, несмотря на это, его все равно применяют, считают это расстояние. Получается некорректно, значит, то есть получаются некие числа, которые, в общем-то, некорректно использовать для каких-то решения каких-то задач. Но тем не менее, это делается, всё равно они используются для решения задач эти числа.
Ну, то есть производятся некорректные расчёты, короче говоря. Вот о чём речь идёт.

В Эйдосе применяются корректные интегральные критерии, представляющие собой скалярное произведение, которое корректно, независимо от того, ортонормировано пространство или нет. Всё равно оно корректным является.
Более того, оно даже корректно в случае, если оно вообще не плоское, как эвклидово пространство, а искривлённое, к примеру. Как в теории относительности Эйнштейна. И даже если оно топологически не является плоскостью.

Ну здесь надо, конечно, заниматься. И вот в результате мы можем э ввести э данные в систему, и эти данные идентифицировать, решить задачу идентификации, прогнозирования. При этом мы, я показываю, что если мы решали задачу формализации раньше предметной области, то есть создавали модель, а теперь мы эту модель применяем для решения задачи прогнозирования.
И у нас это, в общем-то, неплохо получается.

Вот.

Получились у нас результаты, совпавшие с высокой степенью достоверностью с фактом. То есть модели оказались с высокой достоверностью у нас, с достаточно высокой достоверностью.
Вот я вам показываю, как произошло прогнозирование.

И решаются также другие задачи. То есть мы можем решить задачу э определить обратную по отношению к задаче э прогнозирования или идентификации, мы можем определить, какие значения факторов обуславливают переход объекта моделирования в те или иные будущие состояния. Вот здесь у нас эта задача решена. И также решены другие задачи, которые могут решаться с помощью этих моделей.

Можем определить смысл любого признака, любого значения фактора, как он влияет на объект моделирования, и будет ли это значение фактора влиять так, что он переходит в те или иные будущие состояния, которое, в общем, э описывается, в какой степени этот фактор влияет на переход в те или иные будущие состояния, способствует или препятствует, в какой степени.

Рассчитывается матрица сходства, ребята. И потом э эта матрица сходства классов, то есть будущих состояний, она визуализируется в форме когнитивной диаграммы вот такой и в форме дендрограммы классов. То есть эти вот классы, которые объединяются в кластер на высоком уровне сходства, они обуславливаются сходными факторами. Мы видим конструкт классов, то есть видим, что э есть классы, которые обуславливаются одними факторами, есть которые другими, и между собой они сильно отличаются по системе обуславливающих их факторов эти классы. То есть бывает либо вот это, либо вот это. Одновременно то и то быть не может. На когнитивной диаграмме мы видим эту же самую информацию, но на дендрограмме, мне кажется, что она более детально представлена, более в таком развёрнутом виде. Здесь мы видим в более упрощённом варианте эту информацию. Красными линиями соединены классы, которые сходную имеют систему детерминации, а синими – те, которые различную. И вот мы видим, что у нас все будущие состояния образуют конструкт. И на полюсах конструкта э есть э состояния, образующие два кластера. Между собой на полюсах конструкта они все как-то там сходны. То есть там красные линии соединяют их. А полюса конструкта между собой отличаются существенно. Нет ни одной линии красной, ни одного, ни одной линии сходства между этими противоположными кластерами.

То же самое, аналогичный анализ мы можем провести по значениям факторов. То есть мы можем определить, какие из этих значений факторов оказывают сходное влияние на объект моделирования, какие отличающиеся. Насколько эти факторы сходны и отличаются друг от друга по их смыслу. Вот мы здесь показан фрагмент матрицы сходства значений факторов. Небольшой фрагмент, потому что, видите, здесь вот диагональ идёт вот куда. Потом можете себе представить, куда она там идёт дальше.
Вот.
И э это же матрица сходства тоже может быть визуализирована либо в виде когнитивной диаграммы круговой, либо в виде дендрограммы.
Ну, что касается дендрограммы, при таком достаточно большом количестве значений факторов, она получается большой размерности эта дендрограмма.
И чтобы её отобразить, нужно уже использовать э большее разрешение по вертикали. В системе Эйдос э разрешение по вертикали для дендрограмм может быть до 4К. То есть там 4096 э пиксели, пиксели э получается по оси Y. Что, в общем-то, уже так прилично, я бы сказал, довольно много. Сейчас я вам покажу эту дендрограмму, как она выглядит.

Ну вот, так вот, но тут ещё надо, видимо, увеличить масштаб. Ну вот так вот видно, да, что написано, смотрите. Здесь вот кластеры объединены классы в один кластер, которые очень сходны по системе детерминации. То есть факторы, которые обуславливают переход вот в это состояние, в это 67, 79, они очень сходны. То есть когда наблюдается вот это, то наблюдается и вот это 67 и 79 наблюдаются часто одновременно. И с ними же может наблюдаться восьмидесятый ещё вариант или 106. То есть тоже довольно высокий уровень сходства.

И здесь приведена матрица межкластерных диаграмма межкластерных расстояний, график межкластерных расстояний. Сначала при объединении в кластеры классов, э различие растёт медленно. То есть получаются кластеры, в которых кластеры имеют высокий уровень сходства друг с другом. А потом постепенно эта задача усложняется, и приходится объединять в кластеры те, которые уже не очень похожие объекты. Мы здесь видим в диаграмме кластерного анализа, видим, что уже здесь вот на этих уровнях различий приходится объединять. И потом просто уже сильно отличаются друг от друга. И тоже они объединяются.

Здесь мы видим э часть э модели, отображённую в виде нейронной сети.

Изменилось число публикаций у меня за это время, с того времени, когда я писал этот текст. Изменилось очень существенно.

Почему я здесь упоминаю про публикации, ребята? Потому что вот эти 3D интегральные когнитивные карты я реализовал в начале девяностых годов. С тех пор постоянно пишутся книги по когнитивным технологиям, интеллектуальным технологиям, в которых пишут авторы этих книг, пишут, что предлагаются трёхмерные, 3D когнитивные диаграммы, которые, которых задача программной реализации которых, то есть визуализации этих диаграмм, задача для будущих исследователей, будущих разработчиков и будущих исследователей. И я этим авторам пишу: "Ребята, с начала девяностых годов существует система Эйдос, с девяносто второго года она существует, система Эйдос на IBM совместимых компьютерах, которая выводит эти 3D когнитивные диаграммы. Есть работы девяносто первого года, девяносто третьего, где эти когнитивные диаграммы приведены. Понимаете? О которых пишут, что они будут получены в будущих разработках.
Вот. Есть семьдесят девятый год, восемьдесят первый год, девяносто второй год. Вот. Акты внедрения восемьдесят седьмого года вот в системы, которые реализуют эти все вещи.
Вот. То есть, ну вообще просто. То есть это такое впечатление, что люди вообще не читают моих работ. Будто они и не знают, что вообще я существую и работаю, веду разработки. Я написал огромное количество работ, ребята, 628 работ на данный момент. 38 монографий издано с СБЭНами, понимаете? типографии. Мне их девать некуда, они в гараже там лежат уже в кабинете. Вот.

Вот эти диаграммы, которые в будущем будут разработаны, которые существуют давным-давно. Причём авторы вот этих вот книг, которые не читают мои работы, даже и не подозревают о том, что они существуют давным-давно.
Ну мне лично мне обидно. Я об этом пишу, говорю, как об стенку горох. Будто я это не им пишу, понимаете, непонятно кому.

Ведь я ж пишу это для людей, которые этим интересуются, занимаются.

Ну вот.

Да.

### Заключение и Административная информация

Ну что, ребята, какие у вас есть вопросы? Давайте задавайте вопросы. Попробую на них ответить.
В том числе можно задавать вопросы по разработкам приложений, которыми вы должны подтвердить, что вы приобрели необходимые знания, умения и навыки. А я пишу, что на следующей лекции мы рассматриваем вопрос такой очень интересный - когнитивные функции. Я вам рассказываю, что такое когнитивные функции, привожу примеры когнитивных функций, объясняю, что это даёт.

Так, это у нас тут не тот у нас.

Вот, следующий вопрос 1 10 2 5 9 - когнитивные функции.

Ребят, сегодня я сообщаю вам, что сегодня я получил такую информацию от заведующего кафедры, что на следующей неделе начинается аттестация у нас, аттестационная неделя.
Значит, я хочу напомнить, что у нас есть лекционные занятия и лабораторные, практические занятия. На лекционных занятиях вы получаете знания, а на лабораторных и практических занятиях вы эти знания применяете на практике и приобретаете умение их приобретать, то есть умение их, эти знания применять на практике. А потом отрабатываете это умение до уровня навыков. То есть когда вы уже особо так не задумываетесь, что делать, а когда вы уже знаете, что делать, понимаете, с этими знаниями. И приобретаете этим самым соответствующие компетенции, которые у нас предусмотрены в данной дисциплине.
То есть, э вы должны подтвердить свои знания, умения и навыки для того, чтобы пройти нормально аттестацию. Когда была система занятий очная, когда мы занимались в аудиториях, мы видели, преподаватели видели ваши компьютеры, на которых вы работаете в классах, и видели, какой у вас уровень. И в общении, сейчас вот я рассказываю, вы молчите там, молча сидите. Вот. А, значит, на занятиях очных вы спрашивали что-то, задавали вопросы, мы на них отвечали. Я видел, какие вопросы более разумные, какие менее разумные. На лабораторных и практических занятиях мы вообще просто общались. Вот. И у каждого преподавателя было впечатление, мнение о каждом из учащихся. Сейчас это полностью разорвана связь. Мы меня там видите ещё как-то там на каком-то квадратике маленьком, я вообще вас никого не вижу.
Вот. То есть никакой обратной связи у меня нет о том, как вы выглядите. И слушаете вы меня вообще или нет? У меня нет об этом информации. Я вижу, что вы там соединились. Но соединиться можно и пойти спать, например, понимаете? Или соединиться и идти на кухню пить чай. А я здесь рассказываю, тут что-то такое объясняю. Вы можете вообще на всё это дело просто не обращать никакого внимания, что-то там даже можно звук убрать, чтобы я не мешал вам смотреть телевизор и смотреть себе, допустим.
Вот. Я об этом даже и не узнаю.
Вот. Поэтому, значит, э нужно как-то подтвердить, что вы приобрели знания, умения и навыки. Для этого вам даётся задание вот это, которое я подробно объяснял, в чём смысл этого задания, в том, чтобы вы взяли и сделали собственное приложение. Данные где взять для приложения, как его сделать, как описать это приложение, всё здесь подробно описано. И как это потом будет оцениваться, тоже описано.
Значит, ну сейчас я решил смягчить оценку немножко, потому что я вижу, что вы не успеваете, потому что вы думаете, что на всё времени очень много, на самом деле его мало времени. Вот. И не удаётся за то время, которое есть, вот зарегистрироваться в РИНЦ, например, или в ResearchGate. Не получается. Вот. Единицы только смогли это сделать, там буквально один-два человека. Вот. То есть получается так, что мне придётся аттестовать вас по тому, что вы вообще сделали приложение и описали его. Значит, я вам сообщаю вот сейчас вот на лекции сообщаю, что тем, кто сделает приложение, которое я тему, естественно, должен утвердить этого приложения, и опишет его в формате статьи научной, я их аттестую и поставлю им сам экзамен пятёрку. А вот тем, кто что-нибудь не сделает, ну, допустим, не сделает вообще приложение, ну, естественно, не опишет, тогда это три балла, ребята. Если опишет, то я посмотрю, как описал, может быть, четыре балла. Если всё сделает приложение, но не не описал, допустим, ну тогда четыре, может быть, поставлю.

Но я могу вам сказать, что если вы сделаете приложение, не опишете, то оно не будет ни в Эйдос-облаке его не будет, ни, естественно, ни в ResearchGate, ни в РИНЦе. Если вы же его опишете, и это описание будет более-менее нормальное, похоже вот на эти шаблоны, которые я вам дал для описания, тогда я сам его могу разместить, если вы не способны, так сказать, не зарегистрировались там по тем или иным причинам. Я вас, кстати, ребята, не обвиняю в этом, потому что я знаю, что у вас нет э корпоративных адресов электронной почты, как, скажем, в КГУ. Там студент поступает в университет, ему сразу присваивают э адрес корпоративной электронной почты, сразу же, понимаете? Вот. С первого курса. У нас, значит, э этого нет, к сожалению. То есть в плане информационных технологий мы отстаём очень сильно от других вузов. И это я видел ещё, когда пришёл в аграрный университет, сделал ряд предложений Ивану Тимофеевичу Трубилину, которые начали осуществляться через 8-10 лет и через 12 лет после того, как я их сделал. Причём только частично стали реализовываться эти предложения.
Вот. То есть мы очень сильно отстаём. И вот из-за того, что у вас нет этих адресов, которые подтверждают ваш статус как студентов вуза, вы не можете в ResearchGate зарегистрироваться. То есть вы просто теряете уже многое. Вы теряете возможность бесплатных международных публикаций с присвоением DOI, размещение этих публикаций в РИНЦ. Дело в том, что РИНЦ размещает такие публикации, понимаете, с присвоением DOI, опубликованные в ResearchGate. А вы не можете в нём зарегистрироваться и публиковать свои публикации. То есть в РИНЦ вы могли бы зарегистрироваться, это просто требует времени. Но у вас получается, что как бы и незачем, потому что у вас нет публикаций этих вот ResearchGate. Понимаете? То есть вам нечего в РИНЦ размещать. Тут нужно и то, и то, понимаете? Тут нужно и там, и там зарегистрироваться. После этого вы сможете проводить исследования, так как вот я вас учу в системе, сможете описывать их, как вас я учу вот по шаблонам, сможете размещать, как я вас учу в ResearchGate и РИНЦе, и у вас будут появляться публикации, ребята, которые вам нужны для портфолио и для стипендии, для получения стипендии.

Так что давайте, имейте в виду, что вот в четверг уже, ещё несколько дней, и начнётся аттестация. И вот я вот так вот и сделаю. Кто сделает описание, сделает приложение, сделает описание, то значит, пятёрка. Я сам эти описания размещу и приложение в Эйдос-облаке. Кто не сделает описание – четвёрка, но уже не буду нигде размещать, потому что без описания размещать нет смысла. И кто не сделает приложение, ну тройку поставлю. Ну будет там у нас 70% троек, значит, будет 70% троек. А вот в прошлом году, ребята, когда занимались ваши предшественники по этой же дисциплине, э занимались на первом курсе, то там э процентов 80 студентов э стали авторами публикаций. И я очень много поставил пятёрок, а остальные, кто не стал, они тоже очень хорошо занимались, но они просто не сделали эти описания. Вот. Но в классе работали отлично. И я э поставил всем пятёрки. Это вызвало вопросы у учебного управления. Я тогда, ну там всё это, ну не получило никакого таких последствий не было. Почему? Потому что начался COVID, и про это забыли просто.

Я вам показываю, ребята, э статьи этих студентов. Видите, ребята? Это статьи вот этих студентов, ваших предшественников. Они не смогли зарегистрироваться тоже тогда в ResearchGate и РИНЦе, я просто их разместил там. Но для того, чтобы их разместить, мне пришлось э, ну размещать публикации в ResearchGate и РИНЦе может только автор. Поэтому пришлось мне быть соавтором этих студентов. А из-за того, что я был их соавтором, получилось у меня большое число публикаций, в которых я ссылаюсь на самого себя, понимаете?
Вот. То есть перешёл, получилось избыток публикаций с самоцитированием, понимаете?
Вот. Потому что во всех этих публикациях там идут ссылки на мои разработки. И они сейчас тоже идут. И все вот тех, которых вы будете делать. Получается, что мне сейчас ни к чему, понимаете, эти публикации. Тем более, что у меня их вполне достаточно.
Вот. У меня сейчас 628 публикаций, 38 монографий, 27 учебных пособий, три с грифами умо и министерств, 291 публикация в изданиях, входящих в перечень ВАК. То есть мне достаточно публикаций. Я не нуждаюсь в том, чтобы ещё э особенно в этом, так сказать, какие-то усилия предпринимать. Но я могу вам оказать такую помощь в размещении, если вы не успеете там зарегистрироваться. И при этом у меня публикаций увеличится количество, а э индекс самоцитируемости тоже увеличится, что мне не очень хотелось бы делать. А вот э число публикаций увеличится – это нормально, это так вот ничего плохого нет. А вот то, что индекс самоцитируемости увеличится – это плохо.
Поэтому я не хотел бы быть у вас там соавтором в этих статьях. Вот. Но если так выйдет, то буду тогда. Вот. Если вы напишете описание, сделаете, и не сможете разместить, ну придётся мне тогда этим быть соавтором. Но это у меня есть полное моральное право, потому что я разработал все эти технологии и учу вас их применять для научных исследований, а также учу вас описывать эти результаты исследований и размещать публикации на международных площадках и на нашей российской самой крупной библиографической базе РИНЦ, Российский индекс научного цитирования.

Ну так вот примерно. Такая вот ситуация, такова жизнь. Теперь давайте вопросы, пожалуйста, ребята.
Значит, я думаю, что для многих из вас сейчас была новая информация, потому что для тех, у кого была лабораторная работа, я это говорил, а для для остальных, это вот сегодняшняя информация, сегодня она мне поступила, и я вам сегодня её сообщаю, чтобы вы знали, что на следующей неделе будет аттестация. Ну и думайте, надо оно вам, не надо, это уже ваши вопросы.
Вот. Ну вас э руководство университета жалеет. Вот. И, в общем-то, и я тоже э сочувствую вам, что у вас образование получается не совсем то, какое хотелось бы. Вот, такое вот, такой форме. Ну, конечно, эта форма, она уступает всё-таки личному общению, безусловно. Вот. Ну лекции получаются более-менее, я так считаю. А вот лабораторные работы, практические, они получаются более такими ущербными, чем реальные, в реальном, так сказать, контакте, в реальном времени. Ну, получается что? Что вы будете как бы иметь э то есть вам сложнее получить тот уровень квалификации, который получался у студентов, когда они очно учились. Естественно, на Западе есть исследования уже, которые говорят о том, что учащиеся, которые учились в период Ковида, будут меньше зарабатывать. То есть им меньше будут платить просто-напросто. Уже это исследование опубликовано. Вот. Ну руководство вас жалеет, и мы тоже жалеем преподаватели. Поэтому я откровенно говорю, двоек мы вам не поставим. В этом плане вы, так сказать, будете учиться и дальше всё, значит, но это не означает, что не надо ничего делать. То есть я вас призываю к тому, чтобы всё-таки вы предпринимали определённые усилия и подтверждали, что вы действительно учились, получали эти знания, умения, навыки. И вашим подтверждением является вот по нашей дисциплине является подтверждением работа, выполненная вами, наподобие лабораторных работ, описанная так, как вот в научных статьях это описывается.
Вот. Вы на это совершенно, я могу сказать так, уверен абсолютно, что вы на это способны, потому что ваши предшественники прекрасно с этим справились с этой задачей, прекрасно просто. Я даже скажу больше того, они настолько здорово с этим справились, что люди, преподаватели глазам своим не поверили.
Вы понимаете, э, на курсе, вот у вас сколько там человек на курсе там на вашем? Вот. 70 публикаций за семестр, понимаете? 70! Ну вы что, это вообще, понимаете? Вот. Там может быть их не 70, конечно, я их не считал, честно сказать. Но очень много, понимаете?
Вы видели? Можно посчитать, конечно.
То есть ребята прошлого вот курса прекрасно поработали, и всех удивили. И меня они удивили, знаете чем? Тем, что они довольно-таки, ну, скажем так, вполне готовы к таким исследованиям, к таким работам. И они могут и освоить и системы, и технологии, и описать всё, и опубликовать. Вот они как бы уже на первом курсе даже проявляли такую довольно-таки неплохую квалификацию. Что вообще меня поразило, честно сказать, я не ожидал даже. То есть ребята, которые вполне, так сказать, могут всё это въезжать во всё это, делать это всё, дальше двигаться.

У меня есть вопрос. Пожалуйста.
А получается, нужно сдать до четверга Эйдос приложение и описание к нему, да?
А почему до четверга?
Ну а до какого? Ну вы говорите до аттестации или как?
Ну, аттестация будет на основе этого проходить. То есть вы представили, значит, тогда, соответственно, я сказал. Если есть предложение, описание - пятёрка, есть только приложение, описания нет - четвёрка, ничего нет - тройка. Всё. Так вот.
Это будет тогда, когда будет аттестация. Когда крайний срок аттестации? Ну, в принципе, следующая неделя – это аттестационная неделя. То есть, наверное, ну, наверное, пятница – крайний срок аттестации, мне кажется. Если так. Но я точно не знаю, надо ещё уточнить будет. На у нас группа есть на кафедре, значит, такая Ватсаповская. Я там спрошу, когда срок аттестации, сообщу вам крайний срок.
Аттестация проставляется в системе компьютерной. Вот, в той же самой, где посещение вот там же. Там же и ведомости есть. Значит, проблема в чём, ребята? У вас всё-таки не пять человек. И это потребует времени. Вот ещё в чём дело. То есть я вот не могу в пятницу после обеда взять и провести аттестацию, я не успею просто. Понимаете? Это требует времени. То есть, скорее всего, вот когда будут занятия у нас, вот тогда и будет аттестация. Вы на занятиях будете мне эти данные предоставлять. Я буду смотреть. И вот так вот в таком режиме эта аттестация пройдёт. И отмечать буду.
То есть реально это вот так возможно. По-другому это просто трудно вообразить, как это может быть по-другому. Только так, наверное, и может быть.
Ну, всего самого-самого хорошего, ребята. До свидания.
До свидания. До свидания. До свидания. До свидания.