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

***Российская Федерация***

**35 Меры достоверности моделей 2020-11-07**

**Заголовок:** Оценка достоверности моделей в системе Эйдос: F-мера Ван Ризбергена и ее обобщения проф.Е.В.Луценко

**Резюме**

Лекция посвящена методам оценки достоверности моделей в интеллектуальной онлайн-среде Эйдос. Занятие проводит профессор Луценко Е. В.

**1. Введение и постановка задачи**
Рассматривается F-мера Ван Ризбергена как классический критерий достоверности моделей и ее обобщения, разработанные профессором Луценко. Подчеркивается важность оценки достоверности для практического применения моделей.

**2. Настройка системы Эйдос**
Демонстрируется запуск системы Эйдос, создание модели и установка учебного приложения (3.03) для дальнейшей работы.

**3. Классическая F-мера Ван Ризбергена и ее недостатки**

* **Определение:** F-мера основана на суммировании числа истинно положительных (TP), истинно отрицательных (TN), ложноположительных (FP) и ложноотрицательных (FN) решений.
* **Проблемы:**
	+ **Многоклассовость:** Классическая F-мера предполагает принадлежность объекта только к одному классу, тогда как в Эйдос объект может относиться ко многим классам одновременно.
	+ **Нечеткость:** F-мера не учитывает степень уверенности системы в принятом решении (уровни сходства), суммируя лишь бинарные исходы (0 или 1). Это занижает оценку достоверности, особенно при низких уровнях сходства ошибочных решений.
	+ **Зависимость от объема выборки (Инвариантность):** F-мера использует абсолютные количества решений (сумматоры TP, TN, FP, FN), которые линейно растут с увеличением объема выборки. Это делает меру нестабильной и зависимой от размера данных, особенно на малых выборках, где наблюдаются сильные колебания.

**4. Обобщения F-меры (Луценко)**

* **Нечеткое многоклассовое обобщение (L1):** Учитывает степень уверенности (уровень сходства) при суммировании решений, что дает более справедливую и высокую оценку достоверности. Решает проблему нечеткости и многоклассовости.
* **Инвариантное нечеткое многоклассовое обобщение (L2):** Дополнительно к L1, эта мера инвариантна относительно объема обучающей выборки. Она быстро стабилизируется (уже при ~500 объектах) и не подвержена сильным колебаниям на малых выборках, в отличие от классической F-меры.

**5. Практическое значение оценки достоверности**

* **Интерпретация результатов:** Частотное распределение уровней сходства для истинных и ложных решений позволяет установить порог для отсечения недостоверных результатов.
* **Принятие решений:** Достоверность модели критична для ее использования в прогнозировании и управлении. Недостоверная модель приведет к ошибочным прогнозам и неэффективным управляющим воздействиям.
* **Аналогия:** Проводится аналогия с риском применения непроверенных социальных теорий (например, марксизма) в больших масштабах без предварительной оценки их "достоверности" и последствий.

**6. Интегральные критерии сходства в Эйдос**

* Система Эйдос использует интегральные критерии ("Сумма знаний") для сравнения объекта с обобщенными образами классов. Эти критерии основаны на скалярном произведении векторов, описывающих объект и класс.
* **Свойства:** Этот подход не является метрическим в классическом смысле (не связан со свойствами пространства типа ортонормированности или кривизны), но эффективен для сравнения. Он аналогичен разложению сигнала по базисным функциям (как в рядах Фурье, Тейлора и др.), где образы классов выступают в роли базисных функций.
* **Подавление шума:** Благодаря суммированию информации по многим признакам (аналогично теореме Колмогорова о сложности), случайный шум во входных данных подавляется, и оценка сходства становится более робастной.

**7. Выводы**
Обобщенные критерии достоверности (L1, L2) в системе Эйдос предоставляют более адекватную и стабильную оценку качества моделей по сравнению с классической F-мерой, учитывая многоклассовость, нечеткость решений и инвариантность к объему выборки. Это позволяет надежнее интерпретировать результаты распознавания и принимать обоснованные решения на основе моделей.

**Детальная расшифровка текста**

**# 1. Введение и постановка задачи**

(0:04) Игра длится с 15:00 до 16:30
(0:08) по дисциплине Персональная интеллектуальная онлайн-среда Эйдос.

(0:15) И у нас тема занятия
(0:18) меры достоверности моделей,
(0:21) F-мера Ван Ризбергена
(0:23) и её обобщение, которое разработано профессором Луценко.
(0:28) Два обобщения.
(0:29) И
(0:30) информация такая: занятие ведёт
(0:32) профессор Луценко Евгений Венеаминович.

**(Техническая настройка, не относится к содержанию)**
(0:35) (пауза, клики мыши)
(0:53) Попробуй, Александр, этот вариант.
(0:56) Если получится, это будет хорошо.
(0:58) Я попробовал.
(1:01) Да ты что? FTP заработал?
(1:02) Не, не работает.
(1:04) А, не работает, да?
(1:05) Не работает, та же самая ошибка.
(1:07) Ага, понятно.
(1:09) Ну, не знаю.
(1:13) Есть, конечно, разработчик этой библиотеки, можно ему написать.
(1:17) Ладно. Значит,

**# 2. Настройка системы Эйдос**

(1:22) создаём мы приложение
(1:24) в системе.
(1:27) Сейчас мы это сделаем.
(1:29) И потом будем рассматривать этот вопрос.
(1:32) Создаём приложение. Бегло я это всё без комментариев.
(1:36) Вот.
(1:37) Учебное приложение 3.03.
(1:39) Устанавливаем его.
(1:42) (пауза, клики мыши)
(1:48) Вот. И создаём модели.
(1:52) И проверяем их на достоверность
(1:55) путём решения задачи идентификации
(1:58) объектов обучающей выборки. Копируем его в распознаваемую и идентифицируем их.
(2:06) И, значит, извините, сейчас ещё раз
(2:11) здесь вот надо 30 написать.
(2:15) (пауза, звуки клавиатуры)

**# 3. Классическая F-мера Ван Ризбергена и ее недостатки**

(2:31) И переходим в режим 3-4,
(2:34) который предназначен как раз для оценки достоверности модели.
(2:38) И что мы здесь видим в этом режиме? Что у нас есть
(2:41) три
(2:42) разных критерия достоверности
(2:44) в системе Эйдос используется.
(2:46) Это F-мера Ван Ризбергена
(2:49) и её нечёткое мультиклассовое обобщение
(2:52) мера L1
(2:54) Луценко. И L2,
(2:57) которая ещё является и
(2:59) инвариантным относительно объёма данных обобщением F-меры Ван Ризбергена.
(3:05) То есть нечётким, инвариантным, мультиклассовым
(3:09) обобщением. Инвариантным относительно объёма
(3:12) исходных данных.

**# 4. Доступ к Help и ресурсам**

(3:15) И мы можем посмотреть
(3:21) на Help
(3:22) этого режима, где описывается F-мера Ван Ризбергена и упоминается про то, как она
(3:29) там обобщена.
(3:31) И что нас, собственно говоря, не устраивало там в этой мере.
(3:35) Этот Help я посылаю в чат.
(3:38) Ещё раз.
(3:40) Это я бегло, так сказать, ту
(3:43) информацию, которая на прошлом занятии
(3:46) напоминаю.
(3:49) (пауза)
(3:51) И что нас, собственно говоря, не устраивает, показываю. Значит, здесь вот у нас
(3:56) в этом режиме, в хелпе,
(3:58) в режиме 3-4, есть внизу ссылочка на статью.
(4:03) Инвариантно относительно объёма данных. Вот эту статью сейчас я открою
(4:07) и по ней расскажу вам
(4:10) об этом.
(4:12) (пауза, клики мыши, звуки клавиатуры)
(4:31) Посылаю в чат ссылочку на статью.
(4:34) Открываю статью.
(4:38) Статья семнадцатого года,
(4:41) в которой сначала рассматривается
(4:46) классическая традиционная
(4:49) мера достоверности модели F-мера Ван Ризбергена.
(4:54) Потом она записывается из программистских обозначений переводится в математические. Почему? Потому что потом обобщение мультиклассовое, нечёткое, инвариантное, они все потом будут получаться путём э-э
(5:09) ну, математического обобщения, а не программистского, алгоритмического.

**# 5. Критика Классической F-Меры**

## 5.1. Проблема Многоклассовости

(5:15) И говорится о том, что эта классическая мера имеет проблемы определённые,
(5:19) которые мешают её применению.
(5:22) Первая проблема - это проблема мультиклассовости.
(5:25) В чём смысл этой проблемы? Дело в том, что Ван Ризберген предполагает,
(5:30) что один объект выборки относится к одному классу. Он не может относиться к нескольким классам одновременно.
(5:38) В системе Эйдос объект, один объект обучающей выборки
(5:42) может относиться к десяткам, сотням
(5:45) классов одновременно.
(5:48) То есть, ну, точно не к одному, двум, трём, четырём,
(5:50) сотням и тысячам я не сказал, ну до полутора тысяч классов.
(5:55) До полутора тысяч.
(5:57) Это связано чисто с программной реализацией, ограничение. То есть, в принципе, теоретически, если взять модель саму, то там вообще нет никаких таких ограничений особых, только ограничения есть у программной реализации различных. Вот у системы Эйдос текущей версии такое ограничение, где-то около 2.200 этих классов может быть. Ну, реально я пробовал решать различные задачи, где очень большое число классов. Ну, до полутора тысяч классов она нормально всё устойчиво делает. А по литературе, если посмотреть, то должна до 2.300 классов работать.

## 5.2. Проблема Нечеткости (Учета Степени Уверенности)

(6:32) Вторая проблема - проблема нечёткости.
(6:35) То есть это означает, что классическую меру надо обобщать, разрабатывать мультиклассовое обобщение.
(6:41) Проблема нечёткости. В чём заключается эта проблема? В том, что э-э Ван Ризберген суммирует единички к значениям сумматоров э-э True Positive, True Negative, False Positive, False Negative, независимо от того, какая степень уверенности системы в тех или иных решениях.
(7:02) Сейчас мы возьмём,
(7:05) сделаем текущим текущую наиболее достоверную модель
(7:10) и решим задачу идентификации
(7:12) наиболее достоверной модели
(7:15) на графическом процессоре. И посмотрим результаты идентификации.
(7:20) У нас получается такая вот ситуация.
(7:25) Смотрите, что некоторые объекты обучающей выборки
(7:29) идентифицируются с классами э-э как на очень высоком уровне сходства, так и с низкими уровнями сходства.
(7:38) Вот, допустим, объект сумка 1
(7:40) похож на стул на 4%.
(7:43) То есть уровень сходства 4 сотых, если брать э-э Ризберген суммирует единички,
(7:49) к истинно положительным, ложноположительным, э-э, значит, э-э, истинно отрицательным и ложноотрицательным решениям.
(7:56) А система Эйдос определяет, что она
(7:59) эта сумка 1 сходна э-э с образом стул, обобщённым образом класса стул на 4%.
(8:08) То есть, вернее, на 4%, 4 сотых.
(8:11) То есть, если он суммирует единичку, а здесь 4 сотых, то я рекомендую суммировать не единички к этим сумматорам, а степень уверенности системы в своём решении,
(8:20) которая меняется от нуля до единицы
(8:25) и имеет знак.
(8:27) Но суммируется всегда модуль к сумматорам.
(8:30) Значит, ну, в данном случае мы вместо единички к ложноположительным решениям, к сумматору ложноположительных решений, просуммируем 4 сотых. Ясное дело, что это существенно повысит достоверность модели
(8:42) в количественном выражении.
(8:44) То есть это не потому, что модель стала лучше, а потому, что мы стали более справедливо её оценивать.
(8:49) Она ошибается на 4%, мы берём на 100% э-э суммируем там стопроцентный сумматор. Это как-то, ну, несправедливо. Вот, значит, эта несправедливость исправлена
(9:00) в этом отнечётком обобщении.

## 5.3. Проблема Инвариантности (Зависимость от Объема Выборки)

(9:06) Следующая проблема э-э F-меры Ван Ризбергена состоит в том, что в ней, как вы видели, используются сумматоры.
(9:14) То есть это абсолютные величины.
(9:19) Вот, а абсолютные величины, они имеют такое свойство, что они линейно растут с увеличением объёма выборки.
(9:27) Ну я сейчас вам нарисую, как они растут,
(9:31) чтобы было, так сказать, понагляднее это.
(9:35) Если взять у нас
(9:36) оси координат,
(9:38) здесь изобразить у нас
(9:40) объём выборки
(9:44) по оси X.
(9:47) Объём
(9:48) выборки.
(9:54) Вот, а здесь
(9:58) количество
(10:01) решений.
(10:04) И у нас есть разные решения: истинно положительные, истинно отрицательные.
(10:13) Быстрее всего растут
(10:15) истинно отрицательные решения.
(10:18) Немножко медленнее истинно положительные.
(10:21) И потом идут
(10:25) ложноположительные, а потом ложноотрицательные.
(10:32) Значит, быстрее всего растёт число
(10:35) истинно
(10:37) отрицательных решений.
(10:40) Почему? Потому что это решение легче всего осуществлять.
(10:45) Вот. Потом истинно
(10:48) положительных, True Positive.
(10:53) Потом ложно
(10:59) Господи, False Positive.
(11:04) А потом False... Медленнее всего растут
(11:07) ложно
(11:09) отрицательные решения.
(11:13) Их количество.
(11:22) Но как-то это не вдохновляет, ребят, потому что я вам скажу, э-э, рассчитывается ж э-э, полнота, точность и полнота модели.
(11:35) И при этом используются значения этих сумматоров, видите, True Positive, True Negative. Возникает естественный вопрос: а как ведут себя э-э точность и полнота модели, сама достоверность модели э при увеличении объёма выборки,
(11:51) которая использовалась при тестировании
(11:54) модели?
(11:57) И в этой статье
(12:02) этот вопрос исследован.
(12:06) Ну, сначала здесь классически описывается выражение
(12:09) для меры, вот Ризбергена, потом вводятся математические обозначения.
(12:16) И потом с ними дальше идёт работа.
(12:23) Сейчас я опишу суть э-э того, как решена задача.
(12:27) Каким образом она решена.

**# 6. Обобщения F-меры и их поведение**

(12:32) Значит, вот смотрим,
(12:36) как распределены,
(12:40) э-э, как как изменяется
(12:42) достоверность модели
(12:45) при
(12:46) увеличении объёма выборки.
(12:48) Если мы её измеряем с помощью различных мер.
(12:54) Если мы измеряем с помощью F-меры, тогда вот эта жёлтая линия.
(12:59) Вот как меняется достоверность модели.
(13:07) То есть она идёт низко
(13:10) и при малых выборках,
(13:12) где-то примерно до 300, э-э напоминает затухающие колебания периодическое, э-э, то есть сначала очень большая амплитуда колебаний у классической F-меры, потом эта амплитуда уменьшается, и она сходится к некоторой такой плавной плавной кривой, которая стабилизируется при объёме выборки где-то около 2.500.
(13:37) Ну, надо было, конечно, посчитать до 3.000, тогда было бы более убедительно. Но, может, когда-нибудь сделаю. Короче говоря, э-э где-то когда у нас выборка больше 2.500 объектов, тогда происходит такая, ну, скажем так, э-э значение F-мер Ван Ризбергена асимптотически стремится к некоторому пределу и приближается к нему при выборках, начиная где-то с 2.500, фактически она уже э-э асимптотически приближается к этому пределу,
(14:06) который можно считать э-э предельным предельным случае, в предельном переходе э от частот к вероятностям, можно считать, что это и есть истинное значение э меры Ван Ризбергена.
(14:26) Для нечёткого мультиклассового обобщения мы видим, что кривая вот эта сиреневая идёт гораздо выше.
(14:33) Ну, примерно на 1/10, при том, что там у нас максимум единица. Ну, можно сказать так, на 10% выше она идёт. Почему? Потому что более справедливо оцениваются ошибки системы.
(14:46) Если она ошибается, то суммируется к сумматорам э-э False не единицы, а вот эта вот сама ошибка суммируется.
(14:56) 4%, там, 7%, сколько там э-э насколько она уверена в этом решении, эта уверенность и суммируется.
(15:05) Ну а поскольку система в ошибках не уверена,
(15:08) этот, это я уже говорил, что чем выше уровень сходства, тем больше доля истинных решений. Получается, что если мы посмотрим э уровни сходства истинных и ложных решений, то получается, что для истинных решений высокие значения уровней сходства характерны, а для ложных решений низкие уровни сходства. Поэтому получается, что когда мы нечёткое обобщение F-меры разрабатываем, применяем, то у нас получается, что ложные решения с низким уровнем сходства, модуля сходства, и получается, что они вносят меньший вклад.
(15:45) Ну, то есть получается модель более достоверной.
(15:48) Формальная оценка достоверности модели выше. И мы видим, что при малых выборках э нечёткое мультиклассовое обобщение меры Ван Ризбергена не испытывает таких страшных колебаний, как э-э классическая мера Ван Ризбергена.
(16:06) Значит, э что касается классической меры, то вообще можно сказать, что из того, как она себя ведёт, можно предположить, сделать вот такое вывод э предварительный, что, наверное, ей вообще нельзя пользоваться при очень малых выборках. То есть, когда выборки 300 и менее объектов, то просто она даёт, ну почти что фонарные результаты какие-то, понимаете?
(16:28) Но если эти результаты так вот посмотреть на них, так вот глазами, как говорится, попробовать понять, что они означают, то я могу вам сказать, что нужно брать максимальные значения, а не минимальные достоверности. Вот если вы продолжим мысленно эту кривульку, ну можно это сделать не мысленно, а тренд нарисовать там, вот, то мы увидим, что максимальные значения ложатся на неё, на её продолжение, а минимальные сильно выпадают.
(16:52) Согласны, да, ребята? То есть получается, что мера Ван Ризбергена при малых выборках сильно занижает достоверность модели.
(17:02) И нужно брать максимальные значения, чтобы она была более-менее адекватной.
(17:08) И смотрим на синюю вот эту линию, график э-э
(17:16) график зависимости от объёма выборки
(17:21) э-э нечёткого мультиклассового обобщения меры Ван Ризбергена,
(17:27) инвариантное относительно объёма выборки. В чём эта инвариантность заключается, ребята?
(17:32) Вообще, я скажу так, что достоверность-то касается модели, а не выборки. Поэтому получается так, что по идее, когда мы измеряем достоверность модели, её зависимость от объёма выборки, то у нас должна получиться некоторая величина, которую вот мы хотим использовать для оценки достоверности модели, которая будет мало зависеть от объёма выборки.
(17:54) Это ж характеристика модели, а не выборки, понимаете?
(17:57) Вот. Поэтому, значит, она должна мало зависеть от объёма выборки. И мы видим, что вот это инвариантное обобщение, которое было предложено, оно при 500 уже объектах обучающей выборки стабилизируется и потом меняется крайне незначительно. И также не подвержено колебаниям
(18:15) при малых объёмах выборки. То есть она быстро сходится к истинному значению уже при 400
(18:23) объектах выборки, ну при 500 вот здесь мы можем видеть. И даёт оценку достоверности значительно более высокую, чем
(18:32) э просто мультиклассовая нечёткая мера. И ещё гораздо более высокую, чем классическая мера Ван Ризбергена. Ну я могу вам так сказать, что это вот, допустим, классическая мера даёт 0,4, чуть пониже, 0,38. А это даёт 0,7.
(18:48) На той вот модели, которую я здесь вот мы исследовали. А, ну, 0,68. То есть 0,38, 0,68. Где-то на три
(18:59) десятых,
(19:02) вот, то есть это на 30%, она даёт более высокую оценку, потому что максимум равен единице.
(19:10) Ну я вам скажу, что это очень существенно.
(19:15) Вот. То есть это существенно меняет вообще выводы о модели.
(19:19) Но эти выводы меняются обоснованно, потому что у нас появляется критерий для оценки достоверности модели.
(19:29) Этот критерий представляет собой частотное распределение
(19:36) числа истинных и ложных решений. И мы видим, что ложные, то есть решения отрицательные, они все истинные, а положительные решения есть и ложные, и истинные. Но ложные при малых уровнях сходства,
(19:49) ниже, ну, если не смотреть на округлённые, на сглаженные кривые, а сами, на сами оригинальные кривые, то где-то примерно 27%.
(19:59) Вот. Выше нету ложных решений положительных.
(20:05) А истинные решения положительные начинаются с 37%. И ниже нет ложных истинных решений. То есть ложных положительных решений нету. Есть только истинные положительные решения при более высоких уровнях сходства.
(20:17) Ну, фактически это значит, что мы можем смотреть на этот вот 30%, ниже есть ложные решения, выше нет.
(20:24) Выше только истинные решения при более высоких уровнях сходства. Это в модели inf3.
(20:31) Поэтому, если мы сейчас возьмём эту модель inf3, сделаем текущей
(20:35) в режиме 5-6
(20:38) и проведём идентификацию объектов выборки в этом, в этой модели текущей
(20:46) и посмотрим на результаты идентификации,
(20:50) то мы не обнаружим ни одного ложного решения с уровнем сходства
(20:54) выше 30%. Вот смотрите, сейчас вот э-э смотрим.
(21:04) Смотрим на интегральный критерий Сумма знаний.
(21:19) Вот. То есть все решения истинные.
(21:34) Выше 30% уровня сходства.
(21:37) Сейчас ещё посмотрим, обновим.
(21:49) То есть такое вот исследование показывает нам, что мы можем э-э разумно и контролируемо оценивать результаты идентификации.

**# 7. Практическое Значение Оценки Достоверности (продолжение)**

(22:01) То есть мы просто знаем, что ложные решения вот такие уровни сходства имеют, истинные такие.
(22:18) Ну вот, собственно говоря, э-э, что я хотел вам рассказать про
(22:25) меры достоверности.
(22:28) То есть мера достоверности даёт нам критерий для того, чтобы интерпретировать результаты задачи решения, результаты решения задачи идентификации.
(22:37) И оценивать вообще достоверность модели,
(22:40) насколько ей можно пользоваться. Ну, я хочу сказать, что если модель достоверна, то тогда результаты идентификации, прогнозирования соответствуют действительности. Принимаемые решения будут осуществляться,
(22:53) то есть, вернее, э-э, выработка решения о факторах управляющих, которые должны перевести
(23:00) объект управления в заданное целевое состояние,
(23:04) будут осуществляться.
(23:06) То есть этот объект так действительно и будет переходить в целевое целевые состояния.
(23:19) Если же модель недостоверна, тогда идентификация будет ошибочная, прогнозирование не будет осуществляться.
(23:29) При принятии решений об управляющих факторах, эти факторы мы когда выработаем с помощью модели, используем их реально для оказания воздействия на объект управления, то он не будет переходить в те целевые состояния, ради которых мы и всё это вырабатывали эти управляющие решения,
(23:45) ради переходов, в которые вырабатывали управляющие решения. То есть получается так, что если модель адекватна, ей можно пользоваться, а если она неадекватна, то нельзя пользоваться.

**# 8. Историческая/Философская Аналогия**

(23:56) Ну и теперь хочу вам сказать, что вот э-э
(23:59) 104 года назад, да? Э-э, была революция. Э, 103 года назад. В семнадцатом году, это ж двадцатый. Вот. Ну, 13, 103 года назад. Была революция. Её одни называют переворотом, другие, ну так традиционно называется Великая Октябрьская социалистическая революция. Вот. И была принята такая идея, э-э, так сказать, ну, реализована на практике эта идея, э-э, э-э, модернизировать общество, произвести глубокую модернизацию общества в соответствии с учением Маркса, Энгельса, Ленина.
(24:41) Вот. Ну, на основе учения Маркса. Потом, когда уже Ленин умер, тогда стали называть Маркса, Энгельса, Ленина это учение. Потом, когда Сталин появился, начали называть его ещё и Сталина. Маркс, Энгельс, Ленин, Сталин. Вот так стали говорить.
(24:56) Вот. Ну, э-э, что я могу сказать? Что теория очень мощная, очень детально разработанная, капитально. Так вот это слово капитал появилось, собственно говоря, стали его в этом смысле использовать, после того, как Маркс назвал свой основной труд Капитал. Капитальная теория, очень основательно разработанная.
(25:18) Значит, эта теория, она была э-э гипотезой до тех пор, пока её не реализовали. Когда её реализовали, она стала теорией.
(25:27) Вот насколько успешно её реализовали? Если считать, что её реализовали успешно, то она стала теорией. Если считать, что не очень успешно её реализовали, тогда, значит, она так и осталась гипотезой.
(25:38) То есть, если какую-то модель реализуют и получается не очень хорошо, значит, это не совсем адекватная модель.
(25:45) То есть она так и остаётся гипотезой, причём неподтверждённой гипотезой. Ну сейчас всё больше людей склоняется к тому, что это была неподтверждённая гипотеза. Когда мы, значит, смотрели наши предки там, ну это скорее уже даже э не дедушки и бабушки, а там прадедушки, прабабушки, даже ещё более раннее поколение, даже для меня, смотрели на эту теорию, то она производила на них впечатление, ну не на них, а вот на руководство страной, на партии, руководство партиями, э что это, в общем-то, очень солидная, хорошо обоснованная теория научная, которой можно пользоваться. То есть она такое, ну, серьёзное впечатление производила на людей. Вот. И были, правда, конечно, люди, которые критически к этому всему относились. Ну они, значит, были репрессированы там, потом пострадали. Значит, ну я хочу сейчас сказать, ну или просто там затаились. Но я хочу сейчас сказать другое, что э-э, наша э-э руководство, э-э, ну, которое, значит, политическое, которое захватило власть в семнадцатом году, оно, значит, э-э, этой теорией руководствовалось, начало её реализовывать. Значит, что здесь интересно вот в том плане, в котором я сейчас вам рассказываю вот то, что касается моделирования, применения моделей. То, что они некую модель, разработанную выдающимся учёным, это никто не отрицает, что он выдающийся учёный, Маркс я имею в виду. Вот, и помощник его тоже довольно выдающийся Фридрих Энгельс. Значит, конечно, они, ну, он разработал выдающуюся теорию, фундаментальную, серьёзную. Вот. Но никто не знал, насколько она правильная, понимаете?
(27:25) И вот представьте себе, что не проверив достоверность этой теории, не проверив правильность этой теории, начали её реализовать на практике. Это было, это было довольно-таки авантюрно
(27:37) вообще-то. Вот, рискованно. И потом мы увидели, что и гражданская война была, и голод был, и причём не просто голод, а голодомор, как в Петербурге, когда фашисты, в Ленинграде, когда фашисты блокировали его, осталась там только дорога жизни, и так далее, и так далее. То есть вообще-то, э возникает вопрос, насколько правомерно применять какие-то, даже очень такие на вид солидные теории, для таких вот масштабных социальных, политических экспериментов, э в масштабе огромной империи, которая занимала там, не знаю, там пятую часть Земли там или, даже сейчас говорят шестую. Тогда она ещё даже была больше, она включала Польшу, Финляндию, Афганистан тоже подчинялся русскому царю. Вот, и так далее, и так далее. То есть получается там огромнейшие территории. Ну то, что сейчас является странами э-э самостоятельными, всё это входило в состав Казахстан, например, страны Азии, Таджикистан, там этот, Узбекистан, все эти вот э-э небольшие азиатские республики, и закавказские республики, они же все входили в состав Российской империи.
(28:54) И вот возникает вопрос: на такой огромной территории стали применять какую-то модель,
(28:59) которую разработали выдающиеся, конечно, учёные. Здесь вопросов не возникает. Но никто эту модель не проверял, понимаете? Вот в чём проблема, вопрос. И получилось там какие-то определённые результаты, там миллионы жертв получились, потом там что-то вроде стало получаться неплохо. Но дело в том, что никто ж не знает, история ж не знает сослагательного наклонения. Никто не знает, что было бы, если бы этого не делали. Может быть, Россия бы стала сверхдержавой гораздо раньше Америки, и сейчас бы так и оставалась бы сверхдержавой.
(29:32) Она-то и стала в виде Советского Союза, но с этими вот, но, понимаете, с каким, какими, так сказать, жертвами.
(29:42) Вот. Ну, правда, никто не знает, что было бы во время Второй мировой войны, если бы была Российская империя, а не Советский Союз. И никто не знает, была бы вообще эта война Вторая мировая. Я имею в виду.
(29:58) Это всё остаётся неизвестно.

**# 9. Интегральные Критерии Сходства в Эйдос (продолжение) и Дополнительные Ресурсы**

(30:05) Значит, я написал статью небольшую, ребят, об этих делах,
(30:09) разместил её в ResearchGate.
(30:12) И сейчас попробую вам ссылочку дать.
(30:17) (звуки клавиатуры)
(30:27) (пауза)
(30:30) (клики мыши)
(30:36) (звуки клавиатуры)
(30:41) Там есть первый файл идёт по-английски, который, на который ссылочка. А потом там есть файлы, которые по-русски.
(30:51) Если вот здесь вот посмотреть, вот, здесь можно выбрать файл, и он будет по-русски.
(31:02) Там моё видение вот этой вот новейшей истории заложено, изложено.
(31:08) У меня видение оригинальное. Я вижу всё это интегрально, как единый процесс, все эти войны, революции,
(31:16) э-э социальные катаклизмы. И от где-то середины, скажем так, XIX века до XXI века всё прослеживается общая логика, которую я попытался изложить.
(31:31) Теперь, ребята, у нас, когда мы модели посмотрели, то есть создали модели и оценили их достоверность, то теперь дальше у нас по порядку нашего изложения идёт решение задач.
(31:47) Сначала давайте рассмотрим решение задач в самом таком простейшем варианте.
(31:54) То есть в варианте вот в этом, каком здесь нарисованы блоки.
(32:01) А потом рассмотрим их взаимосвязи этих задач, это когда мы будем рассматривать задачу принятия решений, и потом мы будем это связывать с другими задачами.
(32:12) Значит, сейчас рассмотрим решение задачи распознавания и идентификации.
(32:19) Значит, как эта задача решается? Значит, мы должны, для того чтобы её решить, посмотреть на содержимое распознаваемой выборки. Что собой представляет описание объектов
(32:31) распознавания, распознаваемой выборки? Это перечисление кодов тех тех признаков, которые есть у этих конкретных объектов.
(32:40) Здесь справа в окошечке перечисление кодов это содержится. А сама сами эти шкалы, коды, из которых там содержатся, они приведены в режиме 2.2 - это описательные шкалы и градации.
(32:56) И нам нужно каким-то образом сравнить объекты, обладающие этими признаками, с обобщающими образами классов.
(33:08) Как это сделать?
(33:10) Вот у нас есть какой-то объект, он имеет определённые признаки. Как его сравнить с обобщающими образами классов?
(33:16) Для этого в системе Эйдос используются два интегральных критерия. Недавно я ещё придумал третий интегральный критерий, но не реализовал и думаю, что и не буду, наверное, реализовывать.
(33:28) Потому что некогда, как вы видели, расписание
(33:34) какое.
(33:37) Вот. И, значит, э-э, что я хочу вам сказать, ребята? Что э-э есть э-э ну что такое вообще интегральный критерий? Это способ заменить набор частных критериев на одно число.
(33:54) Частные критерии мы уже видели, когда рассматривали модели. Мы видели здесь в помощи, что есть разные формулы для расчёта системно-когнитивных моделей, которые разными способами сравнивают условно и безусловно относительные частоты
(34:12) или просто частоту наблюдения признаков э в тех или иных э в объектах тех или иных категорий.
(34:19) И таких способов сравнения, вот я семь здесь написал. Сейчас я ещё несколько знаю, но э-э, там логет, например, ну не буду сейчас о ней говорить, и я их и тогда знал, когда реализовывал эти режимы, но я их почитал про них, и мне они, скажем так, вызвали некоторое сомнение у меня они, вот так скажем.
(34:42) И вот у нас модель, к примеру, есть. Смотрите. Возьму модель inf1.
(34:48) Это мера информации Харкевича. То есть мы знаем из этой модели, да, сейчас слушайте внимательно. Из этой модели мы знаем, какое количество информации о принадлежности или непринадлежности к тому или иному классу содержится в том или ином значении свойства объекта, в том или ином признаке.
(35:08) То есть про каждый признак мы это знаем.
(35:10) И мы видим здесь и положительное количество информации, и отрицательное.
(35:14) Как это понимать?
(35:16) Значит, э-э, если у нас признак наблюдается чаще, чем в среднем, у объектов какой-то категории чаще, чем в среднем по всей выборке, тогда обнаружение этого признака несёт информацию о принадлежности к этой категории. А если этот признак встречается у объектов этой категории, но реже, чем в среднем по всей выборке, тогда этот признак несёт информацию о непринадлежности к этой категории. Но если за хи-квадрат критерий, то это значит у нас теоретическая, то есть фактическая частота может быть больше, чем теоретическая. Тогда, значит, этот признак связан положительно с принадлежностью к этой категории. А если меньше, чем, фактическая меньше, чем теоретическая, значит, он связан, но отрицательно.
(35:57) Значит, приведу вам пример наглядный. Допустим, нам нужно отличить друг от друга студентов и студенток. И у нас есть признаки: э-э длина волос, наличие брюк, наличие телефона. Что мы можем сказать в этом случае об этих э-э неизвестно ком, э то этими признаками обладает.
(36:18) Если признак длинные волосы есть, тогда этот признак встречается в категории студентки намного чаще, чем в среднем. Будем считать, что примерно поровну студентов и студенток. Тогда будем, получится у нас, что этот признак встречается в два раза чаще у студенток, чем в среднем по выборке.
(36:34) Ну, практически в два раза. А если длинные волосы взять, ну тогда э в группе ребят
(36:40) встречаются они или нет? Ребят? Встречаются, только очень редко.
(36:46) Ну там, я не знаю, там доли процента может быть. Понимаете? Значит, что это значит? Это значит, что обнаружение этого признака несёт много информации о том, что это девушка, и много информации о том, что это не парень.
(36:59) Потому что для ребят это очень нехарактерный признак.
(37:03) Теперь смотрим признак брюки, наличие брюк. Он очень мало несёт информации о том, что это парень, и очень мало информации о том, что не девушка. Если волосы много информации несут, то признак брюки мало информации несёт. Почему? Потому что вероятность его встречи или относительная частота, ну будем говорить вероятность, э несколько упрощённо, э в группе девушки э чуть-чуть ниже, чем в среднем по всей выборке. А в группе мальчики чуть-чуть выше, чем в среднем.
(37:33) Ну, допустим, так. В среднем 99% в брюках ходят. Среди ребят 100, а среди девушек 98%.
(37:41) То есть различие есть, но очень незначительное, понимаете?
(37:46) То есть информации в этом признаке будет очень мало о том, что это парни и о том, что это не девушки. И возьмём третий вариант, допустим, наличие мобильного телефона. У него вообще никакой информации не содержится о том, парень это или девушка.
(38:01) Потому что встречается одинаково часто, с одинаковой вероятностью в этих группах и по всей выборке.
(38:08) И вот мы о каждом признаке имеем, то есть знаем, какое количество информации содержится в каждом признаке о принадлежности, непринадлежности объектов с таким признаком к каждому из классов. Вот в модели хи-квадрат есть всегда э-э не заполнены эти вот, то есть всегда заполнены все клеточки, за исключением тех, где не было данных вообще.
(38:29) То есть если вообще не было данных, тогда там пусто. А если какие-то данные были, то там всегда будет посчитано значение слагаемого хи-квадрат.
(38:39) Вот. Значит, нам нужно сравнить таким образом
(38:42) описание конкретного объекта, у которого есть определённые признаки, с описанием образа класса. Образ класса - это колоночка матрицы модели.
(38:53) Как можно сравнить, ребята, два вектора, по сути? Один вектор - это вектор, характеризующий класс, а другой вектор - это вектор, характеризующий распознаваемый объект.
(39:05) Значит, я показываю вам, ребята, интегральный критерий сходства Сумма знаний я его назвал. Ну, может быть, оно было громко сказано, конечно. Можно сказать, что это просто скалярное произведение или межвекторное расстояние.
(39:21) Сейчас, ребята, я вам пришлю книжку, ссылку на книжку. Не ссылку, а вернее, просто её название. Ну вы легко найдёте её в интернете.
(39:35) (клики мыши)
(39:37) Значит, э-э
(39:41) Эта книжка не так давно вышла.
(39:59) Я, значит, мне описана вот эта мера, э-э, которую я использую э-э с восемьдесят первого года.
(40:09) Всё в семьдесят девятом году придумал. Первые модели делал в восемьдесят первом году. То есть это было на 30 лет раньше, чем написана эта книжка. Но дело в том, что я работал в геофизике и занимался системами геофизической разведки. Я знал, как обрабатывается статистическая информация при геофизической разведке.
(40:28) И там как раз некоторые моменты я использовал в системе Эйдос.
(40:32) В частности, значит, там э-э насчёт матрицы частот, что она вычислялась именно абсолютные частоты, а не относительные. А потом только относительные всегда, когда нужно, тогда они рассчитываются. Почему? Потому что если мы сразу будем считать относительные в процессе вот получения информации, то у нас э-э будет большая потеря информации.
(40:55) Ну я сейчас не буду детальнее объяснять, но в общем, это неправильно просто.
(41:01) Вот. А, значит, тут реализовано всё так очень корректно с точки зрения вот этой этого опыта, который я приобрёл, когда пришёл из армии и работал в Министерстве, ну, в тутторском бюро Министерства геологии,
(41:15) в геофизическом тутторском бюро.
(41:17) Вот. И там очень высокий уровень был программирования и математического обеспечения.
(41:22) И кое-чему там научился, мне кажется.
(41:25) Вот. Так вот, ребята, когда мы берём два вектора каких-то и хотим сравнить, насколько они сходны, то есть мера сходства, которая не связана никоим образом э со свойствами пространства.
(41:36) Это очень удивительная мера. Вообще она, знаете, всё гениальное просто. Она вообще простая сама по себе, но очень удивительная, необычная, имеет возможности. Значит, это скалярное произведение. В координатной форме оно вот так выглядит. То есть координаты векторов перемножаются и складываются.
(41:55) Значит, мы могли бы, конечно, использовав логику Иданелс, взять просто и сравнить, то есть просуммировать только количество информации только в тех признаках, которые реально встречаются у объектов.
(42:08) Вот если признак встречается у объекта, который идентифицируется, то суммируем количество информации, содержащееся в этом признаке о принадлежности к некоторому классу. А если не наблюдается этот признак, то не суммировать. Значит, нужно сделать ветвление, цикл по признакам, и там поставить if then. И, значит, если признак есть, тогда суммировать, а иначе перескакивать на следующую строчку команды, короче говоря.
(42:34) Ну, в общем, можно так сделать, да. Но тогда мы не запишем это в виде формулы.
(42:41) То есть это будет алгоритмическое решение. А формула - это более высокий уровень формализации.
(42:49) И, конечно, формула выражается в форме алгоритма, но не всякий алгоритм выражается в виде формулы. Но этот алгоритм простецкий, он выражается в виде формулы, а именно вот такой.
(42:59) Значит, мы там, когда мы сравниваем условия, есть признак или нет, мы должны иметь массив, э, который будет содержать информацию, есть признак или нет. Этот массив - это массив L, описывающий сам объект, э, идентифицируемый объект.
(43:15) Если у него итый признак наблюдается, то этот элемент массива, соответствующий ему, имеет значение единица. А если не наблюдается - нолик. Вот мы сравниваем. Если единичка, тогда, значит, мы суммируем это количество информации. Если нолик - не суммируем. А можем сделать так: мы можем просто умножить координаты вектора класса,
(43:36) вот это житая колонка, итая строчка, просто можем умножить либо на ноль, либо на один. Если на ноль, то это ясное дело, мы, значит, не суммируем это количество информации в этом признаке, потому что его нет. А если один, то суммируем.
(43:51) Ну вот и всё, собственно говоря. То есть это скалярное произведение, оно с точки зрения геометрии представляет собой просто косинус угла между двумя векторами.
(44:01) Если вектора параллельны, тогда, значит, э равно единице скалярное произведение. Если же они взаимно перпендикулярны, ортонормированы, говорят ещё, то тогда нулю.
(44:14) То есть вектора совершенно не похожи.
(44:19) И в векторной форме это записано выше, чем через сумму.
(44:25) Значит, это интегральный критерий Сумма знаний. Какими интересными свойствами, какими интересными свойствами он обладает? Во-первых, он является корректным для неортонормированных пространств. Во-вторых, он вообще является неметрическим критерием сходства.
(44:40) Метрика - это способ измерения расстояния в каком-либо пространстве. Ну, например, в пространстве Евклида можно измерять э-э расстояние, используя теорему Пифагора.
(44:51) Если у нас есть координаты точек, то корень квадратный из э-э суммы квадратов разности координат представляет собой расстояние между двумя точками. Но это только в том случае, если это пространство Евклидово, плоское.
(45:07) И если оно э-э ортонормировано. Это и есть именно Евклидово пространство - плоское, ортонормированное пространство.
(45:16) То есть система координат в этом пространстве декартова - это взаимно перпендикулярные оси, там две или три, смотря какое размерность пространства.
(45:26) Ну или там 100, если оно большой размерности, там у нас-то большая размерность пространства. Сколько шкала описательная, столько и размерность.
(45:37) А у нас эта мера не связана ни с понятием, представлением об ортонормированности пространств, ни э даже с представлением о его плоскости, кривизне, об отсутствии кривизны, ни даже с тем, является ли оно топологически плоским или нет, топологически эквивалентным Евклидову пространству, или оно, допустим, является тором каким-нибудь, или бутылкой Клейна, там, или ещё чем оно там является, листом Мёбиуса. Понимаете? То есть двумерным там или трёхмерным. То есть двумерная поверхность, свёрнутая определённым образом, превращается в лист Мёбиуса.
(46:22) Получается трёхмерное пространство на самом деле.
(46:26) Или дробной размерности.

(46:30) (кашель) Так. Что-то такое?
(46:37) Пятая пара.
(46:42) Простительно на пятой паре немножко откашляться.
(46:51) Следующее интересное свойство этого интегрального критерия. Значит, есть понятие шума. Есть определение шума. Шум - это такой сигнал, у которого э элементы этого сигнала никак не связаны с друг, одни элементы никак не связаны с другими элементами. То есть значение этого сигнала в любой момент времени никак не зависит от других значений этого сигнала в другие моменты времени.
(47:21) Как определяется, зависят эти фрагменты друг от друга или нет? Это определяется путём вычисления корреляции между ними или скалярного произведения.
(47:31) То есть берём два вектора,
(47:34) координатами которых являются значения этого случайного сигнала. Один вектор отражает один его фрагмент этого сигнала, другой - такой же по величине фрагмент, другой этого сигнала. И вычисляем э-э скалярное произведение этих векторов. Получаем некую величину, которая вот такой же формулой выражается. То есть L и Т - это один фрагмент случайного сигнала, а L и Т житое - это другой.
(48:01) Они равны по числу координат,
(48:05) то есть по длине.
(48:08) И дальше слушайте внимательно, предельная теорема. Значит, при увеличении длины этих векторов L и Т и L и Т житое, то есть при увеличении числа М,
(48:19) вот этого,
(48:21) э-э, по которому до которого индекс идёт э-э и, то есть это индекс, соответствующий признаку. То есть при неограниченном увеличении числа признаков модели, э-э, скалярное произведение вот этих вот любых фрагментов сигнала стремится к нулю.
(48:38) Понимаете? Что интересно? То есть это определение шума.
(48:47) Теперь слушайте внимательно. Любой сигнал, любая информация о любом объекте, описанном в нашей матрице исходных данных, всегда э содержится, содержит две компоненты. Одна компонента отражает истинную природу этого объекта, а другая отражает шум,
(49:08) связанный со случайным воздействием большого числа независимых друг от друга факторов, аддитивно влияющих на объект моделирования. То есть влияние этих факторов совместное равно сумме влияния по отдельности.
(49:22) Приведу вам пример. Допустим, есть куст винограда
(49:26) какого-то сорта. Или лоза какая-то есть. На ней есть много листьев. Берём мы эти листья, срываем и сканируем.
(49:34) Получаются разные изображения. Вопрос возникает такой: почему они разные? У них же одинаковый геном у этих изображений. У этих листьев, вернее, не геном не у изображений, а у листьев, естественно. Геном-то одинаковый, а форма листа определяется сортом, геномом этого сорта. Почему же они разные? Ответ очень простой, потому что на них действуют случайные факторы окружающей среды.
(49:59) Которые можно рассмотреть как шум, который накладывается на истинную форму листа, которая определяется геномом.
(50:10) Вот. Теперь возникает вопрос такой: как эту истинную форму выделить?
(50:15) Значит, я могу вам сказать, что поскольку шум имеет такое свойство, что стремится к нулю при увеличении числа примеров, то мы можем просто что сделать? Взять эти листья
(50:27) э-э и э-э в полярной системе координат их э-э описать их, как их форму. Найти центр тяжести изображения - это элементарно, справа и слева одинаковое число пикселей. И сверху и снизу одинаковое число пикселей. Получается пересечение вертикальной и горизонтальной линии, получается координата центра тяжести. Потом относительно него двигаемся до точки, где максимальная э-э контрастность яркостная и цветовая - это граница листа, это его контур. И дальше, ну, естественно, угол меняется с определённой частотой, на определённую величину, скажем, там на градус, например. Получаем некий массив.
(51:09) Если взять массивы, соответствующие всех листьев куста и сложить их, ну, с нормировать их так, чтобы они были ориентированы одинаково. А это можно сделать автоматически, э-э, нормировку эту, поворот листа. Для этого просто нужно э взять, скажем так, один лист какой-нибудь, отсканированный, да, и вот таким образом оцифрованный. А потом взять следующий. А дальше слушайте внимательно, тоже оцифрованный. И сдвигать так, э-э, массив соответствующий второму листа, листу, относительно первого массива, чтобы можно, значит, вычислять каждый раз корреляцию при каждом сдвиге. Сначала нулевой сдвиг, потом на один отсчёт, потом на два отсчёта, и считать корреляцию. И когда у нас э-э корреляция будет максимальна, это означает, что мы эти листы совместили по фазе, по повороту
(51:59) в системе координат. Потому что смещение на одно э-э число, на один отсчёт, соответствует повороту на тот угол, который у нас был между этими отсчётами. Ну, допустим, если мы отчёты снимали через 1°, то смещение на один отсчёт, допустим, этого массива второго листа относительно первого, соответствует повороту второго листа на 1° относительно первого. Уловили, ребята, нет?
(52:26) Вообще, вы понимаете, о чём я говорю?
(52:34) Ой, ну ладно. Ну примерно, да, Александр, примерно.
(52:38) Так вот получается очень... А?
(52:39) Более-менее. Слава Богу. Ну вы ж математики, вы должны понимать.
(52:46) Поэтому я немножко там и добавляю этих терминов математических в изложение. Я так другим не рассказываю, другим по-другому, более упрощённо рассказываю.
(52:55) Вот. Так вот получается что? Что мы найдём там такое смещение по углу этих листов относительно друг друга, вот так я показываю руками поворот, когда они совпадут наилучшим образом.
(53:09) И после этого мы можем считать сумму по отсчётам, по вертикали вот так вот сумму считать. То же самое касается третьего листа, четвёртого листа и так далее. В результате мы получим некоторую кривулечку, в которой сигнал будет равен полезный сигнал, о котором определяется геномом, будет равен э-э сумме, ну, суммарному количеству этих сигналов, суммарному количеству этих листьев. Вот, а шум уменьшится в корень из числа этих листьев.
(53:41) То есть получится, что отношение сигнал/шум вырастет очень существенно. И мы можем таким образом подавить шум и выявить истинную форму этого листа.
(53:51) То есть этот вот этот интегральный критерий, он, поскольку он является определением белого шума, сам по себе, вот сама формула это является его определением, то это означает, что если у нас в исходных данных есть шум, то он будет подавлен как на этапе формирования матрицы частот, так и на этапе решения задачи идентификации.
(54:12) Потому что при э-э большом числе признаков, э случайная информация, которая содержится в модели, она будет подавлена этим вот интегральным критерием.
(54:24) То есть даст вклад всё больше стремящийся к нулю при увеличении э-э объёма выборки и размерности модели.
(54:34) Теперь дальше очень интересная информация, ребята. Я вам сейчас её сообщу эту интересную информацию в такой форме.

**(Завершение лекции, не относится к содержанию)**
(55:33) И недавно я опубликовал статью, которая является половиной вот этого материала, размещённого в ResearchGate. А вторую половину в этом месяце хочу опубликовать.
(55:58) Так. Сейчас я вам дам задание
(56:01) эту статью посмотреть.
(56:05) У неё резко возрастёт рейтинг тогда.
(56:08) Вот смотрите, вот ссылочка, да? Вот давайте нажимайте на эту ссылочку.
(56:13) И читайте эту статью.
(56:24) О, видите, как хорошо! Сразу рейтинг стал был 28, а стал 31.
(56:30) Вот. И смотрим, что здесь написано. Здесь написана сама теорема Колмогорова, некоторое её обсуждение, её смысла. Э-э, про Арнольда не упоминается, потому что теорема Арнольда, она была доказана позже, чем теорема Колмогорова, и она, м-м, скажем так, у теоремы Колмогорова более общий характер.
(56:55) Значит, здесь я её упрощаю эту теорему Колмогорова.
(57:01) Упрощаю.
(57:02) И привожу к такому виду, который соответствует... Ну, кстати, то, что я её упрощаю, я в этом не оригинален. Если почитать про неё в Википедии, в других каких-то источниках, то вы обнаружите, что есть очень много различных вариантов упрощения теоремы Колмогорова. Она очень такая универсальная, очень общий вид имеет. Ну он прямо сознательно к этому стремился, когда её формулировал. Поэтому так оно и, естественно, у него и получилось. Но я могу вам сказать, что вот это упрощение этой теоремы Колмогорова, которое у меня получилось, формула пятая вот эта вот, она, э, по сути дела, представляет собой общую запись для всех рядов, для всех разложений функций в ряды, какие только вообще могут быть. И хочу вам сказать, что реально э реально были разработаны определённые э-э
(57:58) Сейчас я попробую найти.
(58:05) Э-э известно, известные функции разложения э-э функций для разложения каких-то функций математических в ряды.
(58:16) Первым вообще ряд написал Исаак Ньютон. Этот ряд называется бином Ньютона. Потом появились ряды Тейлора, разложение по степенным функциям, которые в Экселе вот реализовано. Э-э, многочлена одной степени там, ну там только до шестой, по-моему, степени. Ряд Макларена, ряд Фурье широко известен, ряд Лагранжа, там интерполяционные эти вот слагаемые путём интерполяции получались. Э-э, полиномы Чебышева замечательные, которые э при аппроксимации, вот ряд Тейлора, он может пройти очень далеко от э линии ломанной, соединяющей точки. Но пройдёт через эти точки, но будет сильно отклоняться от них. А вот ряд Чебышева, полином Чебышева, он не будет отклоняться от этих вот точек больше, чем на некоторую величину, которая там определяется. Ряд Ларана, вот, а также полиномы Лежандра, Лагера, Эрмита, Бесселя, по спецфункциям. Мы когда учился в университете, мы проходили эти эти ряды, почти все изучали их, и даже ряды по спецфункциям изучали. Вот. И могу вам что сказать? Что все эти ряды,
(59:32) они все могут быть записаны одной простой формулой: некоторая функция умноженная на некоторый коэффициент, который тоже является функцией.
(59:44) Вот. Но этот коэффициент является по смыслу коэффициентом корреляции между э-э функцией, которая разлагается в ряд, и э-э слагаемым, которое там записывается.
(59:59) Значит, ну, представьте себе, я сейчас, может быть, несколько так заумно сказал. Ну представьте себе ряд Фурье, например. Значит, что он собой представляет? Это это сумма синусов и косинусов различных частот, умноженная на определённые коэффициенты, которые называются коэффициенты Фурье. Если мы посмотрим на коэффициенты Фурье, на их математическую форму, то она очень напоминает э-э корреляцию, скалярное произведение. То есть, по сути дела, о чём идёт речь? О том, что исходная функция сравнивается с этими слагаемыми ряда Фурье, сравнивается с каждым из слагаемых. И коэффициент корреляции является э-э коэффициентом при этом слагаемом. В результате что у нас получается? Что исходная функция является суммой всех этих слагаемых ряда Фурье, каждый со своим весом в этой сумме, это называется взвешенная суперпозиция, э, то есть взвешенной суммой, где вот этот коэффициент взвешивания является сходством между исходной функцией и функцией вот этого ряда, э-э, синусом или косинусом определённой частоты.
(1:06:07) Вот такая вот мысль. То есть это вот и есть разложение спектр, прямое преобразование Фурье.
(1:14:00) Что делает система Эйдос? Она разрабатывает вот эти базисные функции.
(1:17:00) Это, ребята, я вам очень советую почитать, это даёт глубокое представление о том, что такое распознавание в системе Эйдос.
(1:20:00) Значит, э по сути дела, э-э, система Эйдос создаёт образы базисных функций, по которым происходит разложение в ряд. И потом определяет, насколько э-э, функция, описывающая объект, функция состояния объекта, сходна с обобщёнными образами функций классов. И этот коэффициент сходства - это и есть э-э интегральный критерий. Второй интегральный критерий, который используется в системе Эйдос, он очень похож на первый. И мы видим по форме математической, что это коэффициент корреляции Пирсона. Но если бы сравним с первым интегральным критерием, то мы видим, что он эта форма корреляции Пирсона получается, если стандартизировать вектора классов и э-э состояний объекта. Как их стандартизировать? Взять все координаты сложить, посчитать среднее, вот, и вычесть э-э из значения координаты среднее и разделить на среднеквадратичное отклонение, которое тоже посчитать. У нас получается стандартизация, обычная стандартизация значений. Можно также использовать формулу интерполяции, то есть вычесть из текущего значения минимальное значение и разделить на разность максимального и минимального значений. Тоже получается при этом стандартизация, ребята. И получится другая форма интегрального критерия. Это я это как раз говорю о третьем интегральном критерии. Но суть такая же, как и у этого вот корреляции Пирсона. А что получается в результате? Значит, корреляция Пирсона - это тоже скалярное произведение, только стандартизированных векторов классов и э объектов, идентифицируемого.
(1:34:00) Так вот, если мы возьмём э и сложим э функции классов, прямо вот функция есть вот эта колоночка матрицы сходства, то есть, извините, колоночка прямо системно-когнитивной модели. Берём эту функцию оттуда прямо, умножаем на на значение коэффициента корреляции э этой функции с исходной функцией, описывающей класс, то есть описывающей объект, идентифицируемый. То э и сложим также точно и функции другого класса, умножим на коэффициент сходства э с этой функцией объекта, его образа. И вот так вот сложим их все, то у нас получится разложение функции объекта в ряд по образам классов.
(1:17:00) Вот о чём я хотел вам сказать, собственно говоря. То есть в системе Эйдос реализован обобщённый спектральный анализ. Похоже на Фурье-анализ, похоже на разложение в другие ряды, виды рядов, но только в системе Эйдос разложение идёт не по каким-то э функциям э математическим, а по образам классов, которые сформированы на основе примеров конкретных путём их обобщения.
(1:44:00) Значит, эти обобщённые образы классов их можно назвать эйдосами. Почему? Потому что конкретные объекты являются их какими-то примерами, реализациями этих обобщённых образов. То есть есть некий обобщённый образ и есть его конкретные варианты, примеры. У Платона как раз об этом и говорилось, что Эйдос является исходным объектом, который находится в многомерном пространстве. Это он называл это идеей, этот объект, который в наше пространство проектируется в виде различных вариантов реализации. И вот эти все варианты вместе, если их обобщить, то можно опять вернуться к этому Эйдосу. Так вот система Эйдос, она обеспечивает как раз вот эту индукцию, то есть она обобщает конкретные примеры и формирует обобщённый образ. Причём в этом обобщённом образе там уже содержится информация о том, насколько какой признак характерен, конкретный признак характерен или не характерен для обобщённого образа. Ну сейчас я забегаю немножко вперёд, но я вам покажу это.
(1:56:00) Вот, например, режим 448. Мы видим форму SWOT-анализа здесь. Вот, допустим, берём элемент компьютера. Что характерно для элемента компьютера? Вот это для него характерно, это не характерно. В графической форме. Вот это вот наиболее характерный признак элемента компьютера, это наименее характерный из характерных. А это те, которые для него вообще не характерны. То есть несут информацию о том, что это не элемент компьютера.
(2:17:00) Вот. Так что система Эйдос является системой разложения э-э функций объектов и ситуаций в ряд по функциям обобщённых образов классов.
(2:29:00) Но, значит, в математике в рядах есть определённое требование э к этим функциям, по которым происходит разложение в ряд. Это требование ортонормированности, что сами эти функции, по которым происходит разложение в ряд, они между собой не коррелируют. У них взаимная корреляция у всех равна нулю.
(2:48:00) То есть эти функции подобраны таким образом, чтобы у них корреляция любых двух функций друг с другом была равна нулю. А в системе Эйдос это совершенно не так.
(2:58:00) То есть, наоборот, я б сказал, что образы классов друг с другом коррелируют. Можно, конечно, освободиться от тех, которые провести типа э выбор главных компонент, как в факторном анализе, можно освободиться от коррелирующих классов. Вот, но это, я думаю, не совсем правильно было бы, потому что это приведёт к потере информации интересной и полезной, которая есть в модели.

**# 10. Заключение**

(3:08:00) Вот такие вот функции объектов.
(3:11:00) То есть конкретные объекты описывались, потом обобщалось это всё, а потом конкретный объект разлагался в ряд по этим обобщённым образам. Такого, наверное, вам никто не рассказывал, я так подозреваю.
(3:23:00) Вот. Ну что ж, всего самого-самого хорошего вам, ребята. До свидания.
(3:28:00) Спасибо. До следующей, до следующего занятия.