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

***Российская Федерация***

**41 Меры достоверности моделей. 2020-11-07**

Заголовок: Оценка достоверности моделей в системе Эйдос: F-мера Ван Рисбергена и ее обобщения Луценко

Резюме текста

Лекция посвящена методам оценки достоверности моделей, создаваемых в интеллектуальной онлайн-среде "Эйдос". Основное внимание уделяется F-мере Ван Рисбергена, ее ограничениям и предложенным профессором Луценко обобщениям.

1. Введение и постановка задачи:

Лекция рассматривает меры достоверности моделей в системе "Эйдос", в частности F-меру Ван Рисбергена и ее обобщения (L1 и L2), разработанные профессором Луценко.

2. F-мера Ван Рисбергена и ее ограничения:

Классическая F-мера используется для оценки достоверности моделей. Она рассчитывается на основе количества истинно положительных (TP), истинно отрицательных (TN), ложноположительных (FP) и ложноотрицательных (FN) решений при идентификации объектов. Однако F-мера имеет ряд существенных недостатков:

Проблема мультиклассовости: Неприменима, когда один объект может принадлежать нескольким классам одновременно, что характерно для системы "Эйдос".

Проблема нечеткости: Не учитывает степень уверенности системы в принятом решении (уровень сходства), суммируя все случаи как единицы, что несправедливо занижает оценку достоверности при ошибках с низкой уверенностью.

Зависимость от объема выборки (неинвариантность): Значения TP, TN, FP, FN являются абсолютными величинами и линейно растут с увеличением объема выборки, что делает саму F-меру зависимой от размера выборки и нестабильной на малых выборках.

3. Обобщения F-меры (меры L1 и L2 Луценко):

Для преодоления ограничений F-меры предложены два обобщения:

Мера L1 (нечеткое мультиклассовое обобщение): Учитывает степень уверенности системы (уровень сходства от 0 до 1) вместо простого суммирования единиц. Это делает оценку более справедливой и повышает значение достоверности. Мера L1 также решает проблему мультиклассовости.

Мера L2 (нечеткое мультиклассовое инвариантное обобщение): Является дальнейшим развитием L1 и инвариантна относительно объема обучающей выборки. Она стабилизируется значительно быстрее (уже при ~500 объектах) и дает еще более высокую и стабильную оценку достоверности по сравнению с L1 и классической F-мерой.

4. Интегральные критерии сходства в "Эйдос":

Система "Эйдос" использует интегральные критерии для сравнения объекта с образами классов. Основной критерий "Сумма знаний" является вариантом скалярного произведения между вектором, описывающим объект, и вектором, описывающим класс (столбцом системно-когнитивной модели). Этот критерий не является метрикой расстояния в классическом понимании, но обладает свойством подавления шума и корректен для неортонормированных пространств признаков, характерных для системы.

5. Практическая значимость оценки достоверности:

Оценка достоверности модели критически важна. Достоверная модель позволяет делать адекватные прогнозы и принимать эффективные управляющие решения. Использование недостоверной, непроверенной модели (проводится аналогия с непроверенным применением марксистской теории в России) может привести к катастрофическим последствиям. Система "Эйдос" предоставляет инструменты для количественной оценки достоверности разработанных моделей.

6. Связь с математическими рядами:

Подход "Эйдос" можно рассматривать как обобщенный спектральный анализ, где функция, описывающая объект, разлагается в ряд по базисным функциям, которыми служат обобщенные образы классов (Эйдосы), сформированные на основе эмпирических данных, в отличие от классических математических рядов (Фурье, Тейлора и др.), использующих предопределенные математические функции.

Детальная расшифровка текста

1. Введение и постановка задачи

Игра длится с 15:00 до 16:30 по дисциплине "Персональная интеллектуальная онлайн-среда Эйдос".

И у нас тема занятия – меры достоверности моделей: F-мера Ван Рисбергена и ее обобщения, которые разработаны профессором Луценко. Два обобщения.

И информация такая: занятие ведет профессор Луценко Евгений Вениаминович.

2. Демонстрация работы в системе "Эйдос"

2.1. Запуск системы и создание модели

Значит, дальше мы что делаем? Мы запускаем систему, создаем модель. Попробуй, Александр, этот вариант. Если получится, это будет хорошо.

Я попробовал.

Ну ты что, FTP заработал?

Нет, не работает.

А, не работает, да?

Не работает, та же самая ошибка.

Ага, понятно. Ну, не знаю. Есть, конечно, разработчик этой библиотеки, можно ему написать. Ладно.

2.2. Создание и установка приложения

Значит, создаем мы приложение в системе. Сейчас мы это сделаем. И потом будем рассматривать этот вопрос. Создаем приложение. Бегло я это все без комментариев. Вот. Учебное приложение 303. Устанавливаем его.

2.3. Создание и проверка моделей

Вот. И создаем модели. И проверяем их на достоверность путем решения задачи идентификации объектов обучающей выборки. Копируем его в распознаваемую и идентифицируем их.

2.4. Переход в режим оценки достоверности

И, значит, извините, сейчас еще раз. Вот здесь вот надо 30 написать.

И переходим в режим 3-4, который предназначен как раз для оценки достоверности модели.

3. F-мера Ван Рисбергена и ее ограничения

3.1. Введение в F-меру и ее аналоги в "Эйдос"

И что мы здесь видим в этом режиме? Что у нас есть три разных критерия достоверности в системе Эйдос используется. Это F-мера Ван Рисбергена и ее нечеткое мультиклассовое обобщение мера L1 Луценко и L2, которая еще является и инвариантным относительно объема данных обобщением F-меры Ван Рисбергена. То есть нечетким, инвариантным, мультиклассовым обобщением. Инвариантным относительно объема исходных данных.

3.2. Справка по F-мере

И мы можем посмотреть на Help этого режима, где описывается F-мера Ван Рисбергена и упоминается про то, как она обобщена. И что нас, собственно говоря, не устраивало там в этой мере. Этот Help я посылаю в чат еще раз. Это я бегло, так сказать, ту информацию, которая на прошлом занятии, напоминаю.

3.3. Ограничения F-меры

3.3.1. Проблема мультиклассовости

И что нас, собственно говоря, не устраивает, показываю. Значит, здесь вот у нас в этом режиме, в хелпе, в режиме 3-4, есть внизу ссылочка на статью. Инвариантно относительно объемов данных. Вот эту статью сейчас я открою и по ней расскажу вам об этом.

Посылаю в чат ссылочку на статью. Открываю статью. Статья семнадцатого года, в которой сначала рассматривается классическая традиционная мера достоверности модели F-мера Ван Рисбергена. Потом она записывается из программистских обозначений переводится в математические. Почему? Потому что потом обобщения мультиклассовые, нечеткие, инвариантные, они все потом будут получаться путем математического обобщения, а не программистского, алгоритмического.

И говорится о том, что эта классическая мера имеет проблемы определенные, которые мешают ее применению. Первая проблема – это проблема мультиклассовости. В чем смысл этой проблемы? Дело в том, что Ван Рисберген предполагает, что один объект выборки относится к одному классу, он не может относиться к нескольким классам одновременно. В системе Эйдос объект, один объект обучающей выборки может относиться к десяткам, сотням классов одновременно. То есть ну точно не к одному, двум, трем, четырем, сотням и тысячам я не сказал, но до полутора тысяч классов. До полутора тысяч. Это связано с чисто программной реализацией, ограничением. 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 2200 этих классов может быть. Ну, реально я пробовал решать различные задачи, где очень большое число классов. Ну до полутора тысяч классов она нормально все устойчиво делает. А по литературе если посмотреть, то должна до 2300 классов работать.

3.3.2. Проблема нечеткости (учета степени уверенности)

Вторая проблема – проблема нечеткости. То есть это означает, что классическую меру надо обобщать, разрабатывать мультиклассовое обобщение. Проблема нечеткости. В чем заключается эта проблема? В том, что Ван Рисберген суммирует единички к значениям сумматоров True Positive, True Negative, False Positive, False Negative, независимо от того, какая степень уверенности системы в тех или иных решениях.

Сейчас мы возьмем, сделаем текущим текущую наиболее достоверную модель и решим задачу идентификации наиболее достоверной модели на графическом процессоре. И посмотрим результаты идентификации. У нас получается такая вот ситуация. Смотрите, что некоторые объекты обучающей выборки идентифицируются с классами как на очень высоком уровне сходства, так и с низкими уровнями сходства. Вот, допустим, объект Сумка 1 похож на стул на 4%. То есть уровень сходства 4 сотых. Если брать Рисберген суммирует единички к истинно положительным, ложноположительным, значит, истинно отрицательным и ложноотрицательным решениям. А система Эйдос определяет, что она эта сумка 1 сходна с образом стул, обобщенным образом класса стул, на 4%. То есть, вернее, на 4%, 4 сотых. То есть если он суммирует единичку, а здесь 4 сотых, то я рекомендую суммировать не единички к этим сумматорам, а степень уверенности системы в своем решении, которая меняется от нуля до единицы и имеет знак. Но суммируется всегда модуль к сумматорам. 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 в количественном выражении. То есть это не потому, что модель стала лучше, а потому, что мы стали более справедливо ее оценивать. Она ошибается на 4%, мы берем на 100% суммируем там стопроцентных сумматоров. Это как-то ну несправедливо. Вот, значит, эта несправедливость исправлена в этом вот нечетком обобщении.

3.3.3. Проблема зависимости от объема выборки (неинвариантность)

Следующая проблема F-меры Ван Рисбергена состоит в том, что в ней, как вы видели, используются сумматоры. То есть это абсолютные величины. Вот, а абсолютные величины, они имеют такое свойство, что они линейно растут с увеличением объема выборки. Ну я сейчас вам нарисую, как они растут, чтобы было, так сказать, наглядно это.

То есть если взять у нас оси координат, здесь изобразить у нас объем выборки по оси X. Объем выборки. Вот, а здесь количество решений.

И у нас есть разные решения: истинно положительные, истинно отрицательные. Быстрее всего растут истинно отрицательные решения. Немножко медленнее – истинно положительные. И потом идут ложноположительные, а потом ложноотрицательные.

Значит, быстрее всего растет число истинно отрицательных решений. Почему? Потому что это решение легче всего осуществлять. Вот. Потом истинно положительных, True Positive. Потом ложно положительных, False Positive. А потом False Negative. Медленнее всего растут ложноотрицательные решения. И их количество.

Ну, как-то это не вдохновляет, ребят, потому что я вам скажу, рассчитывается ж точность и полнота модели. И при этом используются значения этих сумматоров, видите, True Positive, True Negative. Возникает естественный вопрос: а как ведут себя точность и полнота модели, сама достоверность модели при увеличении объема выборки, которая использовалась при тестировании модели?

4. Обобщения F-меры (меры L1 и L2 Луценко)

4.1. Исследование зависимости мер достоверности от объема выборки

И в этой статье этот вопрос исследован. Ну сначала здесь классически описывается выражение для меры Ван Рисбергена, потом вводятся математические обозначения, и потом с ними дальше идет работа.

Сейчас я опишу суть того, как решена задача, каким образом она решена.

Значит, вот смотрим, как распределены, как изменяется достоверность модели при увеличении объема выборки, если мы ее измеряем с помощью различных мер. Если мы измеряем с помощью F-меры, тогда вот эта желтая линия. Вот как меняется достоверность модели. То есть она идет низко и при малых выборках, где-то примерно до 300, напоминает затухающие колебания периодическое, то есть сначала очень большая амплитуда колебаний у классической меры, потом эта амплитуда уменьшается, и она сходится к некоторой такой плавной, плавной кривой, которая стабилизируется при объеме выборки где-то около 2500. Ну, надо было, конечно, посчитать до 3000, тогда было бы более убедительно. Ну, может, когда-нибудь сделаю. Короче говоря, где-то когда у нас выборка больше 2500 объектов, тогда происходит такая, ну, скажем так, значение F-мер Ван Рисбергена асимптотически стремится к некоторому пределу и приближается к нему при выборках, начиная где-то с 2500, фактически она уже асимптотически приближается к этому пределу, который можно считать предельным в предельном случае, в предельном переходе от частот к вероятностям, можно считать, что это и есть истинное значение меры Ван Рисбергена.

4.2. Мера L1 (нечеткое мультиклассовое обобщение)

Для нечеткого мультиклассового обобщения мы видим, что кривая вот эта сиреневая идет гораздо выше. Ну, примерно на 1/10, при том, что там у нас максимум единица. Ну, можно сказать так, на 10% выше она идет. Почему? Потому что более справедливо оцениваются ошибки системы. Если она ошибается, то суммируется к сумматорам False не единицы, а вот эта вот сама ошибка суммируется. 4%, там 7%, сколько там, насколько она уверена в этом решении, эта уверенность и суммируется. Ну а поскольку система в ошибках не уверена, это я уже говорил, что чем выше уровень сходства, тем больше доля истинных решений. Получается, что если мы посмотрим уровни сходства истинных и ложных решений, то получается, что для истинных решений высокие значения уровня сходства характерны, а для ложных решений – низкие уровни сходства. Поэтому получается, что когда мы нечеткое 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. Ну то есть получается модель более достоверной. Формальная оценка достоверности модели выше. И мы видим, что при малых выборках нечеткое мультиклассовое обобщение меры Ван Рисбергена не испытывает таких страшных колебаний, как классическая мера Ван Рисбергена.

Значит, что касается классической меры, то вообще можно сказать, что из того, как она себя ведет, можно предположить, сделать вот такое вывод предварительный, что, наверное, ей вообще нельзя пользоваться при очень малых выборках. То есть когда выборки 300 и менее объектов, то просто она дает, ну почти что фонарные результаты какие-то, понимаете? 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 мы 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. Согласны, да, ребята? То есть получается, что мера Ван Рисбергена при малых выборках сильно занижает достоверность модели. Вот. И нужно брать максимальные значения, чтобы она была более-менее адекватной.

4.3. Мера L2 (нечеткое мультиклассовое инвариантное обобщение)

И смотрим на синюю вот эту линию, график зависимости от объема выборки нечеткого мультиклассового обобщения меры Ван Рисбергена, инвариантное относительно объема выборки. В чем эта инвариантность заключается, ребят? Вообще, я скажу так, что достоверность-то касается модели, а не выборки. Поэтому получается так, что по идее, когда мы измеряем достоверность модели, зависит ее зависимость от объема выборки, то у нас должна получиться некоторая величина, которую мы хотим использовать для оценки достоверности модели, которая будет мало зависеть от объема выборки. Это ж характеристика модели, а не выборки, понимаете? Вот. Поэтому, значит, она должна мало зависеть от объема 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 при малых объемах выборки. То есть она быстро сходится к истинному значению уже при 400 объектах выборки, ну при 500 вот здесь мы можем видеть. И дает оценку достоверности значительно более высокую, чем просто мультиклассовое нечеткое мера. И еще гораздо более высокую, чем классическая мера Ван Рисбергена. Ну я могу вам так сказать, что это вот, допустим, классическая мера дает 0,4, чуть пониже, 0,38. А эта дает 0,7 на той вот модели, которую здесь вот мы исследовали. А, ну 0,68. То есть 0,38, 0,68. Где-то на 3 десятых. Вот, то есть это на 30% она дает более высокую оценку, потому что максимум равен единице. Ну я вам скажу, что это очень существенно. Вот. То есть это существенно меняет вообще выводы о модели. Но эти выводы меняются обоснованно, потому что у нас появляется критерий для оценки достоверности модели.

5. Частотное распределение решений и его использование

Этот критерий представляет собой частотное распределение числа истинных и ложных решений. 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, ниже, ну если не смотреть на округленные, на сглаженные кривые, а сами на сами оригинальные кривые, то где-то примерно 27%. Вот. Выше нету ложных решений положительных. 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. Это в модели Inf3.

Поэтому если мы сейчас возьмем эту модель Inf3, сделаем текущей в режиме 5-6 и проведем идентификацию объектов выборки в этом, в этой модели текущей, и посмотрим на результаты идентификации, то мы не обнаружим ни одного ложного решения с уровнем сходства выше 30%. Вот смотрите, сейчас вот смотрим. Смотрим на интегральный критерий Сумма знаний. То есть все решения истинные выше 30% уровня сходства. Сейчас еще посмотрим, обновим.

То есть такое вот исследование показывает нам, что мы можем разумно и контролируемо оценивать результаты идентификации. То есть мы просто знаем, что ложные решения вот такие уровни сходства имеют, истинные такие.

6. Практическая значимость оценки достоверности и историческая аналогия

Ну вот, собственно говоря, что я хотел вам рассказать про меры достоверности. То есть мера достоверности дает нам критерий для того, чтобы интерпретировать результаты задачи решения, результаты решения задачи идентификации и оценивать вообще достоверность модели, насколько ей можно пользоваться. Ну, я хочу сказать, что если модель достоверна, то тогда результаты идентификации, прогнозирования соответствуют действительности, принимаемые решения будут осуществляться. То есть, вернее, решения о факторах управляющих, которые должны перевести объект управления в заданное целевое состояние, будут осуществляться. То есть этот объект действительно и будет переходить в целевое, в целевые состояния.

Если же модель недостоверна, тогда идентификация будет ошибочная, прогнозирование не будет осуществляться. 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 все это вырабатывали эти управляющие решения, ради перехода в которые вырабатывали управляющие решения. То есть получается так, что если модель адекватна, ей можно пользоваться, а если она неадекватна, то нельзя пользоваться.

Ну и теперь хочу вам сказать, что вот 104 года назад, да, была революция, и 103 года назад, в семнадцатом году, то есть двадцатый. Вот. Ну, 13, 103 года назад была революция. Ее одни называют переворотом, другие вот 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. 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 еще и Сталина. Маркс, Энгельс, Ленин, Сталин – вот так стали говорить.

Вот. Ну, э-э, что я могу сказать? Что теория очень мощная, очень детально разработанная, капитальная, 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.

Значит, эта теория, она была гипотезой до тех пор, пока ее не реализовали. Когда ее реализовали, она стала теорией. Вот насколько успешно ее реализовали? Если считать, что ее реализовали успешно, то она стала теорией. Если считать, что не очень успешно ее реализовали, тогда, значит, она так и осталась гипотезой. То есть если какой-то модель реализуют и получается не очень хорошо, значит, это не совсем адекватная модель. То есть она так и остается гипотезой, причем не подтвержденной гипотезой.

Вот сейчас все больше людей склоняется к тому, что это была неподтвержденная гипотеза. Когда мы, значит, смотрели наши предки там, ну это скорее уже даже не дедушки и бабушки, а там прадедушки, прабабушки, даже еще более раннее поколение, даже для меня, смотрели на эту теорию, то она произвела на них впечатление, ну не на них, а вот на руководство страной, на партии, руководство партиями, что это, в общем-то, очень солидная, хорошо обоснованная теория научная, которой можно пользоваться. То есть она такое, ну, серьезное впечатление производила на людей. Вот. И были, правда, конечно, люди, которые критически к этому всему относились. Но 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 наше руководство, ну, которое, значит, политическое, которое захватило власть в семнадцатом году, оно, значит, этой теорией руководствовалось, начало ее реализовывать.

Значит, что здесь интересно вот в том плане, о котором я сейчас вам рассказываю, вот то, что касается моделирования, применения моделей? То, что они некую модель, разработанную выдающимся ученым, это никто не отрицает, что он выдающийся ученый, Маркс я имею в виду, вот, и помощник его тоже довольно выдающийся Фридрих Энгельс. Значит, конечно, они, ну, он разработал выдающуюся теорию, фундаментальную, серьезную. Вот. Но никто не знал, насколько она правильная, понимаете?

И вот представьте себе, что не проверив достоверность этой теории, не проверив правильность этой теории, начали ее реализовать на практике. Это было это было довольно-таки авантюрно вообще-то. Вот, рискованно. И потом мы увидели, что и гражданская война была, и голод был, причем 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 возникает вопрос, насколько правомерно применять какие-то, даже очень такие на вид солидные теории, для таких вот масштабных социальных, политических экспериментов в масштабе огромной империи, которая занимала там, не знаю, там пятую часть земли там или даже сейчас говорят шестую, тогда она еще 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 самостоятельными, все это входило в состав Казахстан, например, страны Азии, Таджикистан, там этот Узбекистан, все эти вот небольшие азиатские республики и закавказские республики, они же все входили в состав Российской империи.

И вот возникает вопрос: на такой огромной территории стали применять какую-то модель, которую разработали выдающиеся, конечно, ученые, здесь вопросов не возникает. Но никто эту модель не проверял, понимаете, вот в чем проблема, вопрос. И получилось там какие-то определенные 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. Она-то и стала в виде Советского Союза, но с этими вот "но", понимаете, с каким, какими, так сказать, жертвами. Вот. Ну, правда, никто не знает, что было бы во время Второй мировой войны, если бы была Российская империя, а не Советский Союз. Никто не знает, была бы вообще эта война, Вторая мировая. Я имею в виду. Это всё остаётся неизвестным.

7. Интегральные критерии сходства и связь с математическими рядами

Значит, я написал статью небольшую, ребят, об этих делах, разместил ее в ResearchGate. И сейчас попробую вам ссылочку дать.

Вот. Там есть первый файл идет по-английски, который, на который ссылочка. А потом там есть файлы, которые по-русски. Если вот здесь вот посмотреть, вот, здесь можно выбрать файл, и он будет по-русски.

Там мое видение вот этой вот новейшей истории заложено, изложено. Мое видение оригинальное. Я вижу все это интегрально, как единый процесс, все эти войны, революции, э-э, социальные катаклизмы. И от где-то середины, скажем так, XIX века до XXI века все прослеживается общая логика, которую я попытался изложить.

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 идет решение задач. Сначала давайте рассмотрим решение задач в самом таком простейшем варианте. То есть в варианте вот в этом, каком здесь нарисованы блоки. А потом рассмотрим их взаимосвязи этих задач. Это когда мы будем рассматривать задачу принятия решений, и потом мы будем это связывать с другими задачами.

Значит, сейчас рассмотрим решение задачи распознавания и идентификации. Значит, как эта задача решается? Значит, мы должны, для того чтобы ее решить, посмотреть на содержимое распознаваемой выборки. Что собой представляет описание объектов распознаваемой выборки? Это перечисление кодов тех признаков, которые есть у этих конкретных объектов. 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.

И нам нужно каким-то образом сравнить объекты, обладающие этими признаками, с обобщающими образами классов. Как это сделать? Вот у нас есть какой-то объект, он имеет определенные признаки. Как его сравнить с обобщающими образами классов? Для этого в системе Эйдос используются два интегральных критерия. Недавно я еще придумал третий интегральный критерий, но не реализовал и думаю, что и не буду, наверное, реализовывать. Потому что некогда, как вы видели, расписание какое.

Вот. И, значит, что я хочу вам сказать, ребят? Что есть, ну что такое вообще интегральный критерий? Это способ заменить набор частных критериев на одно число. Частные критерии мы уже видели, когда рассматривали модели. Мы видели здесь в помощи, что есть разные формулы для расчета системно-когнитивных моделей, которые разными способами сравнивают условные и безусловные относительные частоты или просто частоты наблюдения признаков в тех или иных, в объектах тех или иных категорий. И таких способов сравнения вот я семь здесь написал. Сейчас я еще несколько знаю, но там логет, например, но не буду сейчас о них 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, так скажем.

И вот у нас модель, к примеру, есть. Смотрите. Возьму модель Inf1. 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. То есть про каждый признак мы это знаем. И мы видим здесь и положительное количество информации, и отрицательное. Как это понимать? Значит, 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 несет 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 несет информацию о непринадлежности к этой категории. Но если с точки зрения хи-квадрат критерия, 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.

Значит, приведу вам пример наглядный. Допустим, нам нужно отличить друг от друга студентов и студенток. И у нас есть признаки: длина волос, наличие брюк, наличие телефона. Что мы можем сказать в этом случае об этих, э-э, ну, неизвестно ком, э-э, кто этими признаками обладает? Если признак длинные волосы есть, тогда этот признак встречается у категории студенток 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. А ну, практически в два раза. А если длинные волосы взять, ну тогда в группе ребят встречаются они или нет, ребят? Встречаются, только очень редко. Ну там, я не знаю, там доли процента, может быть. Понимаете? Значит, что это значит? Это значит, что обнаружение этого признака несет много информации о том, что это девушка, и много информации о том, что это не парень. Потому что для ребят это очень нехарактерный признак.

Теперь смотрим признак брюки, наличие брюк. Он очень мало несет 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 несет. Почему? Потому что вероятность его встречи или относительная частота, ну будем говорить вероятность, несколько упрощенно, в группе девушки чуть-чуть ниже, чем в среднем по всей выборке. А в группе мальчики чуть-чуть выше, чем в среднем. Ну, допустим, так: в среднем 99% в брюках ходят. Среди ребят 100, а среди девушек 98%. То есть различие есть, но очень незначительное, понимаете? То есть информации в этом признаке будет очень мало о том, что это парни и о том, что это не девушки. И возьмем третий вариант, допустим, наличие мобильного телефона. У него вообще никакой информации не содержится о том, парень это или девушка. Потому что встречается одинаково часто, с одинаковой вероятностью в этих группах и по всей выборке.

И вот мы о каждом признаке имеем, то есть знаем, какое количество информации содержится в каждом признаке о принадлежности, непринадлежности объекта с таким признаком к каждому из классов. Вот в модели хи-квадрат есть всегда не заполнены эти вот, то есть всегда заполнены все клеточки, за исключением тех, где не было данных вообще. 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.

Вот. Значит, нам нужно сравнить таким образом описание конкретного объекта, у которого есть определенные признаки, с описанием образа класса. Образ класса – это колоночка матрицы модели. Как можно сравнить, ребят, два вектора, по сути? Один вектор – это вектор, характеризующий класс, а другой вектор, то есть другой вектор – это вектор, характеризующий распознаваемый объект.

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 евклидово расстояние.

Сейчас, ребята, я вам пришлю книжку, ссылку на книжку. Не ссылку, а вернее, просто ее название. Ну вы легко найдете ее в интернете.

Эта книжка не так давно вышла.

Я, значит, мне описана вот эта мера, которую я использую с восемьдесят первого года. Я ее 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. И там как раз некоторые моменты я использовал в системе Эйдос. В частности, значит, там насчет 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 будет большая потеря информации. Ну я сейчас не буду детальнее объяснять, но в общем, это неправильно просто. Вот. А, значит, тут реализовано все так очень корректно с точки зрения вот этой этого опыта, который я приобрел, когда пришел из армии и работал в Министерстве, в Конструкторском бюро Министерства геологии, геофизическом Конструкторском бюро. Вот. И там очень высокий уровень был программирования и математического обеспечения. И кое-чему там научился, мне кажется.

Вот. Так вот, ребята, когда мы берем два вектора каких-то и хотим сравнить, насколько они сходны, то есть мера сходства, которая не связана никоим образом со свойствами пространства. 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.

Значит, мы могли бы, конечно, использовав логику Вида Нелса, взять просто и сравнить, то есть просуммировать только количество информации только в тех признаках, которые реально встречаются у объектов. 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, команду, короче говоря.

Ну, в общем, можно так сделать, да. Но тогда мы не запишем это в виде формулы. То есть это будет алгоритмическое решение. А формула – это более высокий уровень формализации. 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. Значит, мы там, когда мы сравниваем условия, есть признак или нет, мы должны иметь массив, который будет содержать информацию, есть признак или нет. Этот массив – это массив L, описывающий сам объект, идентифицируемый объект. Если у него итый признак наблюдается, то этот элемент массива соответствующий имеет значение единицы. Если не наблюдается – нолик. Вот мы сравниваем: если единичка, тогда, значит, мы суммируем, если нолик – не суммируем. А можем сделать так: мы можем просто умножить координаты вектора класса, 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.

Ну вот и всё, собственно говоря. То есть это скалярное произведение, оно с точки зрения геометрии представляет собой просто косинус угла между двумя векторами. Если вектора параллельны, тогда, значит, равно единице скалярное произведение. Если же они взаимно перпендикулярны, ортонормированы, говорят еще, то тогда нулю. То есть вектора совершенно не похожи.

И в векторной форме это записано выше, чем через сумму. 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. Метрика – это способ измерения расстояния в каком-либо пространстве. Ну, например, в пространстве Евклида можно измерять расстояние, используя теорему Пифагора. Если у нас есть координаты точек, то корень квадратный из суммы квадратов разности координат представляет собой расстояние между двумя точками. Но это только в том случае, если это пространство Евклидово, плоское. И если оно ортонормировано. Это и есть именно Евклидово пространство, плоское, ортонормированное пространство. То есть система координат в этом пространстве декартова – это взаимно перпендикулярные оси, там две или три, смотря какое размерность пространства. Ну или там 100, если оно большой размерности, там у нас-то большая размерность пространства, сколько шкал описательных, столько и размерность.

Вот. А у нас эта мера не связана ни с представлением об ортонормированности пространств, ни даже с представлением о его плоскости, кривизне, об отсутствии кривизны, ни даже с тем, является ли оно топологически плоским или нет, топологически эквивалентным Евклидовому пространству, или оно, допустим, является тором каким-нибудь или бутылкой Клейна там, или еще чем оно там является, листом Мёбиуса. Понимаете? То есть двумерным там или трехмерным. То есть двумерная поверхность, свернутая определенным образом, превращается в лист Мёбиуса. Получается трехмерное пространство на самом деле. Или дробной размерности.

Так. Что-то такое? Пятая пара. Простительно на пятой паре немножко откашляться.

Следующее интересное свойство этого интегрального критерия. Значит, есть понятие шума. Есть определение шума. Шум – это такой сигнал, у которого элементы этого сигнала никак не связаны с 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.

Как определяется, зависят эти фрагменты друг от друга или нет? Это определяется путем вычисления корреляции между ними или скалярного произведения. То есть берем два вектора, 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 скалярное произведение этих векторов. Получаем некую величину, которая вот такой же формулой выражается. То есть L и L жито – это другой. Они равны по числу координат, то есть по длине.

И дальше, слушайте внимательно, предельная теорема. Значит, при увеличении длины этих векторов L и L жито, то есть при увеличении числа М вот этого, по которому, до которого индекс идет и, то есть это индекс, соответствующий признаку. То есть при неограниченном увеличении числа признаков модели, скалярное произведение вот этих вот любых фрагментов сигнала стремится к нулю. Понимаете, что интересно? То есть это определение шума. Сама по себе, вот сама формула – это является его определением.

Теперь, слушайте внимательно. Любой сигнал, любая информация о любом объекте, описанном в нашей матрице исходных данных, всегда содержится, содержит две компоненты. Одна компонента отражает истинную природу этого объекта, а другая отражает шум, 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.

Привожу вам пример. Допустим, есть куст винограда какого-то сорта или лоза какая-то есть. На ней есть много листьев. Берем мы эти листья, срываем и сканируем. Получаются разные изображения. Вопрос возникает такой: почему они разные? У них же одинаковый геном у этих изображений, у этих листьев, вернее. 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. Вот. И которые можно рассмотреть как шум, который накладывается на истинную форму листа, которая определяется геномом.

Вот. Теперь возникает вопрос такой: как эту истинную форму выделить? 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, э-э, и в полярной системе координат их описать их, э-э, 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 контрастность яркостная и цветовая – это граница листа, это его контур. И дальше, ну, естественно, угол меняется с определенной частотой, на определенную величину, скажем, там на градус, например. Получаем некий массив. 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 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 отсчет, потом на два отсчета, и считать корреляцию. И когда у нас корреляция будет максимальна, это означает, что мы эти листы совместили по фазе, по повороту в системе координат. Потому что смещение на одно число, на один отсчет, соответствует повороту на тот угол, который у нас был между этими отсчетами. Ну, допустим, если мы отсчеты снимали через 1°, то смещение на один отсчет, допустим, этого массива второго листа относительно первого, соответствует повороту второго листа на 1° относительно первого. Уловили, ребят, нет? Вообще, вы понимаете, о чем я говорю?

Ой, ну ладно. Ну, примерно, да, Александр, примерно. Так вот получается очень, а? Более-менее. Слава Богу. Ну вы ж математики, вы должны понимать. Поэтому я немножко там и добавляю этих терминов математических в изложении. Я так другим не рассказываю, другим по-другому, более упрощенно рассказываю.

Вот. Так вот получается что? Что мы найдем там такое смещение по углу этих листов относительно друг друга, вот так я показываю руками поворот, когда они совпадут наилучшим образом. И после этого мы можем считать сумму по отсчетам, по вертикали вот так вот сумму считать. То же самое касается третьего листа, четвертого листа и так далее. В результате мы получим некоторую кривулечку, которой сигнал будет равен полезный сигнал, который определяется геномом, будет равен сумме, то есть суммарному количеству этих сигналов, то есть суммарному количеству этих листьев. Вот. А шум уменьшится в корень из числа этих листьев. То есть получится, что отношение сигнал/шум вырастет очень существенно. И мы можем таким образом подавить шум и выявить истинную форму этого листа.

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. Потому что при большом числе признаков случайная информация, которая содержится в модели, она будет подавлена этим вот интегральным критерием. То есть даст вклад всё больше стремящийся к нулю при увеличении э-э, объема выборки и размерности модели.

Теперь дальше очень интересная информация, ребят. Я вам сейчас ее сообщу эту интересную информацию в такой форме.

В форме довольно большой работы моей о теореме Колмогорова пятьдесят седьмого года. Про теорему Колмогорова слышали, ребят, нет, что-нибудь? Честно признавайтесь. Не хотите признаваться, что ли?

И недавно я опубликовал статью, которая является половиной вот этого материала, размещенного в ResearchGate. А вторую половину в этом месяце хочу опубликовать.

Вот. И смотрим, что здесь написано. Здесь написана сама теорема Колмогорова, некоторое ее обсуждение, ее смысла. Про Арнольда не упоминается, потому что теорема Арнольда, она была доказана позже, чем теорема Колмогорова, и она, скажем так, у теоремы Колмогорова более общий характер.

Значит, здесь я ее упрощаю эту теорему Колмогорова. Упрощаю. И привожу к такому виду, который соответствует, ну, кстати, то, что я ее упрощаю, я в этом не оригинален. Если почитать про нее 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 ее формулировал. Поэтому так оно, 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, 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-э, реально были разработаны определенные, сейчас я попробую найти.

Известно, известные функции разложения функций, математических, в ряды. Первым вообще ряд написал Исаак Ньютон. Этот ряд называется бином Ньютона. Потом появились ряды Тейлора, разложение по степенным функциям, которые в Экселе вот реализованы, многочлена одной степени там, ну там только до шестой, по-моему, степени. Ряд Маклорена, ряд Фурье, широко известный, ряд Лагранжа, там интерполяционные эти вот слагаемые путем интерполяции получались. Полиномы Чебышева замечательные, которые при аппроксимации, вот ряд Тейлора, он может пройти очень далеко от линии ломаной, соединяющей точки, но пройдет 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. Я 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, они все могут быть записаны одной простой формулой: некоторая функция умноженная на некоторый коэффициент, который тоже является функцией. Вот. Но этот коэффициент является по смыслу коэффициентом корреляции между функцией, которая разлагается в ряд, и слагаемым, которое там записывается.

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 определенные 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 чем идет 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, то есть взвешенной суммой, где вот этот коэффициент взвешивания является сходством между исходной функцией и функцией вот этого ряда, э-э, синусом или косинусом определенной частоты.

Вот такая вот мысль. То есть это вот и есть разложение спектр, прямое преобразование Фурье.

То есть функция разлагается в ряд по слагаемым этого вот спектра по гармоникам.

Что делает система Эйдос? Она разрабатывает вот эти базисные функции. Это, ребята, я вам очень советую почитать, это дает глубокое представление о том, что такое распознавание в системе Эйдос. Значит, по сути дела, система Эйдос создает образы базисных функций, по которым происходит разложение в ряд. И потом определяет, насколько функция, описывающая объект, функция состояния объекта, сходна с обобщенными образами функций классов. И этот коэффициент сходства – это и есть 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 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 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, ребят. 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.

Так вот, если мы возьмем и сложим функции классов, прямо вот функция есть вот эта колоночка матрицы сходства, то есть, извините, колоночка прямо системно-когнитивной модели. Берем эту функцию оттуда прямо, умножаем на на значение коэффициента корреляции этой функции с исходной функцией, описывающей класс, то есть описывающей объект идентифицируемый, то и сложим также точно и функцию другого класса умножим на коэффициент сходства 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. Вот о чем я хотел вам сказать, собственно говоря. То есть в системе Эйдос реализован обобщенный спектральный анализ, похожий на Фурье-анализ, похожий на разложение в другие ряды, виды рядов. Но только в системе Эйдос разложение идет не по каким-то математическим функциям, а по образам классов, которые сформированы на основе примеров конкретных путем их обобщения.

Значит, эти обобщенные образы классов их можно назвать Эйдосами. Почему? Потому что конкретные объекты являются их какими-то примерами, реализациями этих обобщенных образов. То есть есть некий обобщенный 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 обобщенный образ. Причем в этом обобщенном образе там уже содержится информация о том, насколько какой признак характерен, конкретный признак характерен или не характерен для обобщенного образа. Ну сейчас я забегаю немножко вперед, но я вам покажу это. Вот. Это мы будем еще рассматривать. Вот 448 режим. Мы видим форму свод-анализа здесь. Вот, допустим, берем 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 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.

Вот. Так что система Эйдос является системой разложения функций объектов и ситуаций в ряд по функциям обобщенных образов классов. Но, значит, в математике в рядах есть определенные требования 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. 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. То есть, наоборот, я б сказал, что образы классов друг с другом коррелируют. Можно, конечно, освободиться от тех, которые провести типа 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 приводит к потере информации интересной и полезной, которая есть в модели.

8. Заключение и вопросы

Так. Ну, то есть у нас система функций, по которым разложение в ряд происходит функции объекта, она у нас избыточная. Вот. В математической теории рядов этих система этих функций, она минимальна и достаточна. То есть эти образуют полную систему функций. Вот. А в системе Эйдос эта система функций избыточна. Но отражает интересные аспекты предметной области, которая моделируется.

Теперь, если мы возьмем прямо исходные данные в виде прямо каких-то конкретных э-э, ну в общем, соответствующие гармоникам ряда Фурье, и создадим обобщенные образы, потом будем разлагать эти обобщенные образы по э-э, классам, классы будут представлять собой гармоники, а исходный объект будет описываться с каким-то суммарным сигналом, то система Эйдос просто будет производить разложение ряд Фурье. То есть просто она будет его осуществлять. То есть я хочу сказать, что вот этот аппарат, который там реализован, он в частном случае сходится сводится к тем различным э-э, рядам, которые в математике известны.

Вот. Какие вопросы, ребят? Осталось 3 минуты занятия. Ну так уловили примерно, да, о чем я говорю? То есть если мы, допустим, берем у человека исследуем какую-то модель, которая создана на основе примеров людей. То есть люди, вот студенты группы, например, у них есть различные там пол, возраст, вес, там различные их социальный статус, как они учатся и так далее, и так далее. Мы взяли на основе всего этого создали соответствующие обобщенные образы: обобщенный образ студента, обобщенный образ студентки, обобщенный образ отличника, обобщенный образ студента из Кореновска, обобщенный образ студента из Краснодара. Ну так вот вот взяли, создали все эти обобщенные образы такого-то возраста, такого-то веса там и так далее, и так далее. Поняли, да? Потом берем конкретного студента какого-то. И она говорит: "Это у нас студент, а не студентка, это у нас, скорее всего, из Славянска, это вот он учится вот так и вот так на по таким-то дисциплинам и так далее, и так далее". То есть этот тот конкретный студент является взвешенной суперпозицией обобщенных образов классов, сформированных на основе ряда примеров, с этими вот коэффициентами сходства, которые посчитаны с помощью интегральных критериев. То есть можно представить себе, что каждый человек является суперпозицией некоторых архетипов. Но взвешенной, то есть одни более ярко представлены, другие менее ярко представлены. И вот все вместе они составляют именно этого конкретного специфического человека, который описывается конкретно своей функцией специфической. То есть система Эйдос, она вводит ряд обобщений таких, что мы можем описывать объекты функциями, эти функции можем разлагать в ряд по обобщенным образам классов. Вот вы где-нибудь слышали, что объекты можно описывать функциями? Ну, допустим, в квантовой механике там известно, что элементарные частицы можно описывать волновыми функциями. О том, что студентов можно описывать функциями, вы об этом слышали? Слышали. Где? Ну, в целом как-то за 4 года учебы. То есть об этом, об этом вам говорили, да? Да. Ну что ж, это я считаю круто. Но все равно я думаю, что в том, что я сейчас вам рассказывал, есть какой-то элемент оригинальности, новизны и доведенности до реализации, полноты. То есть это же все доведено до уровня технологии, которую можно взять и посчитать, все посмотреть. То есть это не просто какие-то идеи, а это идеи реализованные. То есть в этом я считаю, что есть определенное достоинство того, что я рассказываю. А вот те, кто вам рассказывал про то, что можно человека описать функциями, они могут этот вид этой функции представить? Как она выглядит? Могут они разложить в ряд эту функцию?

Ладно, ребят, на этом у нас занятие заканчивается. В ряд не Фурье или там какой-то другой, а в ряд по образам классов, которые сформированы на основе примеров. Вот таких вот функций объектов. То есть конкретные объекты описывались, потом обобщалось это все, а потом конкретный объект разлагался в ряд по этим обобщенным образам. Такого, наверное, вам никто не рассказывал, я так подозреваю. Вот. Ну что ж, всего самого-самого хорошего вам, ребят. До свидания. Спасибо. До следующей недели. До следующего занятия.