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

***Российская Федерация***

**102 Теория информации, данные и знания. Лабораторная 4. Достоверность модели**

**Резюме:**

Видео представляет собой запись лабораторного занятия №4 по дисциплине "Теория информации, данные, знания" для группы ИТ2001 от 28 сентября 2020 года, проводимого профессорами Луценко Е.В. и Аршиновым Г.А.

**1. Введение и Постановка Задачи:**
Занятие является продолжением предыдущего (ЛР №3), где была создана модель в системе "Эйдос". Текущая задача – изучить достоверность созданных моделей и научиться решать задачи идентификации, прогнозирования, принятия решений и исследования предметной области с их помощью.

**2. Верификация Моделей и F-мера Ван Рисбергена:**

* **Понятие верификации:** Верификация определяется как проверка или оценка достоверности модели.
* **Метод:** Основной метод верификации в "Эйдос" – решение задачи идентификации объектов обучающей выборки с использованием самой модели.
* **Критерий F-мера:** Для оценки достоверности используется классический F-критерий (F-мера) Ван Рисбергена. Объясняется его расчет на основе четырех типов исходов идентификации:
	+ Истинно положительные (True Positive, TP): Объект правильно отнесен к своему классу.
	+ Истинно отрицательные (True Negative, TN): Объект правильно не отнесен к чужому классу.
	+ Ложно положительные (False Positive, FP): Объект ошибочно отнесен к чужому классу ("ложное срабатывание").
	+ Ложно отрицательные (False Negative, FN): Объект ошибочно не отнесен к своему классу ("пропуск цели").
* **Проблемы классической F-меры:**
	+ **Моноклассовость:** Исходно F-мера разработана для ситуаций, когда объект принадлежит только одному классу, что редко встречается в реальности.
	+ **Зависимость от объема выборки:** Сумматоры TP, TN, FP, FN являются абсолютными частотами, их значения линейно зависят от размера выборки, что делает F-меру неинвариантной к объему данных.
	+ **Бинарность оценки ошибки:** При ошибках (FP, FN) в сумматоры добавляются единицы, не учитывая степень уверенности/неуверенности модели в ошибочном решении.

**3. Модификации F-меры в Системе "Эйдос":**

* **Мультиклассовое обобщение:** Предложен вариант F-меры, работающий для объектов, принадлежащих нескольким классам.
* **Нечеткое обобщение:** Вместо добавления единиц при ошибках, суммируется "степень ошибки" (например, уровень сходства при FP или разница до порога при FN), что делает оценку более справедливой и учитывает неуверенность модели.
* **Инвариантность к объему выборки (L2-мера):** Предложено использовать относительные частоты (вероятности) TP, TN, FP, FN, деленные на объем выборки. Эта мера (L2) быстро стабилизируется с увеличением данных и показывает более высокую достоверность, чем классическая F-мера.

**4. Практическая Работа в "Эйдос":**

* **Запуск системы и модели:** Демонстрируется установка лабораторной работы 3.03, запуск синтеза и верификации моделей (статистических и моделей знаний) на центральном процессоре.
* **Анализ результатов верификации:**
	+ **Графики распределения решений:** Показываются гистограммы распределения истинных и ложных решений в зависимости от уровня сходства. Демонстрируется, что ложные решения концентрируются при низких уровнях сходства, а выше определенного порога (например, 40% для данной модели) ложных решений практически нет.
	+ **Интегральные критерии:** Объясняются два интегральных критерия сходства в "Эйдос": "Сумма знаний" (скалярное произведение вектора объекта и вектора класса, неметрическая мера, устойчивая к шуму) и "Семантический резонанс" (коэффициент корреляции Пирсона между стандартизированными векторами объекта и класса).
	+ **Выбор достоверной модели:** На основе анализа F-меры и ее модификаций (L1, L2) выбирается наиболее достоверная модель для дальнейшей работы (в данном случае модель ИНФ3 по критерию L2).

**5. Решение Задач с Использованием Модели:**

* **Идентификация/Прогнозирование:** Задача идентификации (или прогнозирования при динамической интерпретации) решается путем сравнения объекта с обобщенными образами классов. Результат – ранжированный список классов по убыванию сходства. Интерпретация идентификации как разложения функции объекта в ряд по функциям классов.
* **Принятие решений (Обратная задача):** По заданному целевому состоянию (классу) определяется набор значений факторов, максимально способствующих переходу в это состояние (SWOT-анализ). Решается мгновенно путем сортировки строк матрицы модели.
* **Исследование предметной области:**
	+ **Анализ факторов:** Инвертированные SWOT-диаграммы показывают "семантический потенциал" каждого признака – к каким классам он тяготеет.
	+ **Кластеризация классов:** Когнитивные диаграммы и дендрограммы показывают сходство между самими классами на основе их описаний в модели.

**6. Организационные Моменты:**
Обсуждаются вопросы получения студентами корпоративной почты и регистрации на платформе ResearchGate.

**Детальная Расшифровка Текста:**

**1. Введение и Постановка Задачи**

Здравствуйте, ребята.

Вы должны микрофоном отвечать: "Здравствуйте, здравствуйте".
Здравствуйте.
Здравствуйте.
Здравствуйте, здравствуйте.

Ну что, сегодня у нас 28 сентября 2020 года, понедельник. Первая пара проводится с 8:00 до 9:30. Лабораторная работа номер четыре у группы ИТ-2001 по дисциплине Теория информации, данные, знания. Занятие ведет профессор Луценко Евгений Вениаминович и профессор Аршинов Георгий Александрович.

На прошлом занятии, ребята, мы остановились на том, что мы создали модель, модели в лабораторной работе 3.03. И должны на этом занятии продолжить с того, начать с того, что должны изучить достоверность этих моделей. И потом уже решать различные задачи в этих моделях: идентификации, принятия решений и исследования моделируемой предметной области.

**2. Верификация Моделей и F-мера Ван Рисбергена**

* **Понятие верификации и метод**

Я устанавливаю систему. Ищу, кто-нибудь есть у нас, у кого не стоит система Эйдос на компьютере, не установлена? В чате, если есть, то поставьте минусики там в чате, у кого не стоит система Эйдос. Сейчас вы должны по идее всё, что я делаю, повторять на своих компьютерах, на своих компьютерах. Времени для этого вполне достаточно, так как я не просто что-то делаю, а ещё подробно объясняю, и это занимает много времени. Поэтому у вас вполне достаточно времени, чтобы повторить то, что я делаю.

Значит, сейчас, смотрите, ребята, вот я сейчас запускаю систему Эйдос. Если какие-то есть у вас вопросы, проблемы, вот, то спрашивайте, потому что это лабораторная работа, на ней предполагается, что вы вполне свободно задаёте вопросы мне. То есть можете спрашивать. И на почту можете писать тоже вопросы. Я тоже буду отвечать вам.

Вот так. А, наверху появилась. Не разглядел. Вот, запускаем систему. А вы сейчас должны всё это видеть, что я показываю и повторять.

* **Критерий F-мера**

Осуществляется синтез и верификация модели. Синтез проходит вообще очень быстро, а потом довольно долго, сравнительно, проходит верификация. Почему? Потому что при этом рассчитывается 11 выходных форм по результатам решения задачи идентификации, прогнозирования. Ну, принятия решения, правда, нет. Вот. И верификация у нас начинается с режима... То есть, в общем-то, в основном посвящён режим 3-4. Есть другие режимы, где более подробная информация. Но сейчас мы не будем этим заниматься, стандартном варианте посмотрим. Это режим 3-4.

Прежде всего, что такое верификация, ребят? Верификация - это проверка достоверности модели или оценка достоверности модели. Нажимаем помощь. И здесь идёт объяснение, что такое, что понимается под достоверностью модели. Сразу же вам скажу, что используется F-критерий Ван Рейсбергена для оценки достоверности модели и различные его обобщения, модификации, которые я разработал. Почему я их разработал, сейчас объясню. Если бы можно было не разрабатывать, я бы и не стал, если честно. Какой смысл? А в связи с тем, что такая возникла необходимость, ну это я сделал. Вот статья, ребята, где описано всё, что касается критериев достоверности моделей в системе Эйдос. Я вам советую тех, кто, ну я, конечно, всё это объясняю и рассказываю, но те, кто хочет немножко поглубже познакомиться с этим материалом, я всегда ссылаюсь на статьи и книжки свои, пособия. Даю вам такую возможность.

Значит, представьте себе, что мы с вами играем в такую игру. Вот я вам сообщаю, ребята, вы, наверное, об этом не подозревали, но я вам сообщаю, что я прекрасно могу предсказывать события, просто великолепно. Вот представьте себе, что у вас, ну, практически вообще без ошибочно. Вот у вас есть кубик игральный, и вы сейчас можете даже проэкспериментировать. Вот я вам сейчас спрогнозирую, что у вас выпадет, а вы кинете, и у вас выпадет именно то, что я скажу. Как вам это кажется, интересное? Наверное, даже и не верится. Значит, ну я вам сообщаю, ребята, слушайте внимательно, я делаю прогноз. У вас выпадет либо один, либо два, либо три, либо четыре, либо пять, либо шесть. Ну вы чувствуете, здесь какой-то подвох такой. Вот. Ну могу вам сказать, что если сейчас вы кинете кубик, то у вас действительно что-нибудь из этого выпадет, вы же понимаете прекрасно. Вы говорите: "Да, Евгений Венич, ну вы даёте. Ну это, конечно, здорово вы предсказываете, но дело в том, что какая польза от этого предсказания, да, грубо говоря? Когда вы предсказали все возможные варианты". Больше того, я ведь сказал, что выпадет 1 2 3 4 5 или 6, а выпало только три у кого-то, у кого-то только пять, ну и так далее. Ну и кого-то опять тоже только три выпало. А остальные-то не выпали грани. Говорите: "Да, или, ребят, ну немножко я, как, сплаховал. Ну, конечно, но ведь то, что я сказал, ведь выпало же, правильно?" Говорите: "Да, но при этом вы же сказали, что не выпадет остальное, а оно выпало". То есть получается что? Что много ошибок всё равно в моём предсказании, потому что я должен был предсказать, что не только что выпадет, но и что не выпадет. Вот. И при этом получилось, что в том, что выпадет, я был прав один раз, а остальные были ложноположительные решения, потому что я говорил, что, ну и говорил, что выпадет ещё также 1 2 4 5 и 6, а они не выпали. То есть я совершил ошибки, предсказав то, что не произойдёт, что не произошло. Что произойдёт, я предсказал правильно, а что не произойдёт, очень много ошибок произошло, возникло в моём предсказании. Поэтому я скорректирую свои предсказания и понял, я понял, что лучше всё-таки предсказывать, чего не произойдёт. Вот слушайте новое предсказание, скорректированное с учётом первого, с учётом этих вот ошибок. Там было только одно правильное решение, истинно положительное, и было пять ложноположительных решений. То есть решений, когда я говорил, что они произойдут, эти события, они не произошли. Теперь давайте я предскажу ещё один раз, и будет ошибок намного меньше. Вот слушайте внимательно. Вот сейчас вы будете кидать кубик, и у вас вообще ничего не выпадет. Говорите: "Да ну ладно". Вот сейчас вот я кину и выпадет. Вот смотрите, вот я кинул, у меня выпало, вот уже прямо вот сейчас выпала четвёрка. Говорите: "Да, ну, немножко я ошибся, конечно". поэтому по поводу. Действительно выпала четвёрка. Но ведь, но всё ведь, всё, что я говорил относительно остальных вариантов, оно ж оказалось правильным. Я же говорил, что не выпадет ни один, ни два, ни три, ни пять, ни шесть, да? Ну выпало, ну, люди ж ошибаются, бывает, конечно. Я ошибся, что когда я сказал, что не выпадет четыре, выпало четыре. Но в остальном-то я был прав. 1 2 3 5 и 6 не выпало ж. Если посчитать число ошибок, которые я совершил, предсказав второй раз, то получается, что я совершил ошибок очень мало. То есть практически только одну ошибку я совершил, а всё остальное было правильно. Представляете? То есть когда я предсказывал, что произойдёт, то я много допустил ошибок, предсказав, что что-то произойдёт, а оно не произошло. А когда я предсказывал, что не произойдёт, то тогда у меня будет все совершенно правильные мои были прогнозы, кроме одного, который осуществился. Вот. А как должен идеальный прогноз выглядеть? Он должен выглядеть следующим образом. Я говорю: "Выпадет четыре, а остальные не выпадут". То есть я должен конкретно сказать, что выпадет и что не выпадет. Потом проверяем с фактом и смотрим, совпало. О, ничего себе! Это да. Действительно здорово предсказывает Евгений Вениаминович. Вот. Ну, реально так не бывает. Реально всегда мы предполагаем, что что-то произойдёт, какие-то наиболее вероятные варианты, там один, два, там, три. Вот. И какие-то наименее вероятные варианты мы тоже видим. А остальные мы не там, не очень-то и понятно, может оно быть, может и не быть.

Так вот, когда-то в своё время, ну я так думаю, что это было довольно давно. Ну мы можем попытаться конкретно сказать, когда это было. В общем, Ван Рейсберген предложил этот критерий. ну сейчас в очень таком преклонном возрасте, я не знаю, честно говоря, жив он или нет, потому что давно смотрел там информацию о нём. Значит, он что, собственно говоря, предложил? Он предложил очень убедительный критерий достоверности модели, который заключается в следующем. Он начал сначала сформулировал, зачем вообще нужна модель. Модели, ребята, нужна для решения задач. Без без модели задачи решать невозможно. Какие задачи можно решать с помощью модели, когда она есть? Задачи идентификации, диагностики, классификации, распознавания - это всё синонимы этих слова, которые я сейчас произнёс. Прогнозирование. Идентификация и прогнозирование - это почти синонимы, я потом буду рассказывать подробнее про эти задачи, тогда расскажу. Это практически почти что, и в общем, ну не совсем, чуть-чуть отличается, но очень близки по смыслу. Потом дальше задачи принятия решений или поддержки принятия решений, тоже объясню, в чём разница. И задача исследования моделируемой предметной области путём исследования модели.

Ну как же узнать достоверность модели? Ван Рейсберген предложил это сделать, решая задачу идентификации тех же самых объектов обучающей выборки, которые были, на основе которых была создана модель. То есть, чтобы её решить, для этого нужно скопировать обучающую выборку в распознаваемую и произвести в тёмную распознавание, то есть не подглядывая, к каким классам эти объекты относятся обучающей выборки. А потом взять, проанализировать результаты идентификации и посчитать, сколько было истинно положительных, истинно отрицательных, ложно положительных и ложно отрицательных решений. Эти вот термины, они для вас, возможно, являются новыми, поэтому я сейчас их объясню подробно. Значит, есть решения положительные и отрицательные. Положительные - это о принадлежности объекта к классу. Отрицательные - о непринадлежности. Ван Рейсберген говорит так: "Модель должна хорошо, правильно относить объекты к тем классам, к которым они относятся, и должна правильно не относить объекты к тем классам, к которым они не относятся". Что, в общем-то, совершенно точно то, что он сказал. Так и должно быть. При этом система старается это сделать, но не всегда у неё это получается. То есть она может иногда допустить ошибки и в первом случае, и во втором случае. Что означает система относит класс, то есть объект к классу, к которому он относится? Это означает, что когда она вычисляет... В системе Эйдос, если конкретно, вычисляет суммарное количество информации, которое содержится в признаках объекта о принадлежности этого объекта к какому-то классу, и выбирается класс о принадлежности к которому больше всего информации содержится в признаках объекта. И получается, что информация содержится очень много, и по количеству, то есть и знак этого положительный. То есть содержится много информации о принадлежности, то есть знак информации положительный. Если же в признаках содержится информация о непринадлежности объекта к классу, то знак тогда будет отрицательный. Положительное решение - это решение, когда уровень, то когда сходство объекта с классом положительное. И отрицательное решение - это когда сходство отрицательное, то есть это означает различие, а не сходство уже. Решения бывают, могут быть истинными и ложными. Если объект на самом деле относится к классу, и система его относит к этому классу, то это истинно положительное решение. True Positive. Сумматор предложил Рейсберген. Здесь буковки означают не переменные, а вот две буквы сразу обозначают сумматор ТП, True Positive, как в языке программирования, а не в алгебре. А когда объект не относится к классу, и система его и не относит к нему, тогда это называется True Negative, истинно отрицательное решение. Если, ну и ложноположительное, ложноотрицательное. Если объект не относится к классу, а система его относит, то есть система отнесла объект к классу, а он на самом деле к нему не относится к этому классу, тогда это называется ложноположительное решение. Это называется ложное срабатывание. Это опасный вид ошибки. Представьте себе, что нужно идентифицировать, ну, в боевой обстановке нужно идентифицировать противника и своих, чтобы по своим не стрелять. И летит самолёт, и система идентификации делает, даёт ложноположительное решение. Летит наш самолёт, а система сообщает, что это самолёт противника. Вот. То есть ложноположительное решение. То есть она относит его к категории самолёт противника, а на самом деле он не является самолётом противника. И пуск ракеты сбивают его. То есть это может быть очень-очень плачевные результаты иметь. И ложноотрицательное решение. Когда система не относит объект к классу какому-то, а фактически он к нему относится. Что предложил, собственно говоря, Ван Рейсберген? Он предложил посчитать эти сумматоры и предложил формулы, по которым он, в общем-то, оценивает и точность модели, полнота модели, и достоверность модели. Всё прекрасно было бы, если бы не одно но. В системе Эйдос каждый объект обучающей выборки может относиться ко многим классам. И в жизни тоже так. То есть идея о том, что объект может относиться только к одному классу или не относиться - это идея, она правильная, но очень упрощённая. На самом деле объект может относиться ко многим классам. На самом деле фактически относиться. И ко многим, естественно, не относиться. Вот. Ну, я всегда, когда про это рассказываю, привожу пример себя почему-то, я не знаю, почему так вот мне кажется удобным. Ну вот я, например, кто я такой? Я преподаватель Кубанского государственного аграрного университета. Я профессор, я муж, я отец, я дедушка. Понимаете? Вот. Ещё кто я? А ещё я мужчина, да? А ещё кто? А ещё я россиянин. И так далее, и так далее. Ещё ещё кто? А ещё я человек с высшим образованием. Ну там ещё доктор наук. Ну то есть это я сказал, что я профессор, но ещё ж другие есть признаки. Вот. Короче говоря, мы можем вот так вот найти очень много категорий, к которым я отношусь. Вот. И также любой человек, и вообще любой объект. Берём любой объект, вот, он относится не к одному классу какому-то, категории обобщающей, а ко многим. Вот. Берём какую-нибудь собачку. Она что? Эта собачка домашняя, она там определённой породы, она живёт в определённом доме там, в определённом городе и так далее, и так далее. То есть получается, что она относится ко многим классам. Классам, классу собачек, живущим в этом доме, например, она относится. А таких собачек там пять штук, например, которые в этом доме живут.

* **Модификации F-меры в Системе "Эйдос"**

Так вот, теперь, значит, получается, что нужно критерий Ван Рейсбергена как-то обобщить на мультиклассовый случай. Я это сделал. То есть я разработал, математически описал вот в этой статье мультиклассовое обобщение критерия Ван Рейсбергена. Теперь следующий момент. Вот эти вот счётчики, они же, в них указывается абсолютная частота, то есть количество этих вот случаев. Посчитать их относительную частоту, то есть делить их на количество испытаний, количество вот этих вот решений. Ну то есть получается, что я предложил, грубо говоря, заменить частоты относительные, э, то есть абсолютные частоты относительными, которые стремятся к вероятностям. При увеличении объёма выборки эта относительная частота стремится к вероятности. И потом уже эти вероятности использовать для расчёта меры Ван Рейсбергена. Я это сделал, соответствующие математические выражения вывел, посчитал на примерчике всё это начисленном. И у меня получилось, ребята, такая картина. Здесь вот формулы приводятся где-то там пораньше немножко. Вот они для этого случая. Это нечёткое мультиклассовое обобщение меры Ван Рейсбергена. И получилось то, что и ожидалось. Оказалось, что вот эта вот инвариантная относительно объёма данных, нечёткая мультиклассовая мера Ван Рейсбергена, она очень быстро стабилизируется. Значит, я предложил, собственно, две меры. Одна мера мультиклассовая сразу же, то есть нечёткую мультиклассовую меру предложил. Вот, это L1 назвал её, а L2 - это ещё и инвариантная относительно объёма выборки. Она быстро стабилизируется. Эти меры по величине достоверности выше, чем классическая мера Ван Рейсбергена, F-мера. Почему? Потому что они более справедливые. Суммируются к ошибочным решениям, к сумматорам, учитывающим ошибочные решения, суммируются не единички, а степень ошибки. Вот эта величина ошибки и суммируется. Получается, что на самом деле достоверность модели гораздо выше, чем по мере Ван Рейсбергена, но значительно выше, я бы сказал.

**6. Практическая Работа в "Эйдос"**

* **Запуск системы и модели**

Ну я выбрала вот то, что вы сказали, всё равно не получилось. В режиме 3.5 не пошёл процесс? Нет, не пошёл. Выдалась ошибка, да? Угу. Ну пришлите мне эту ошибку, скриншот. Alt Print Screen нажмите. Сейчас пришлю. Пришлите на почту. Ну сейчас я не буду разбираться, потому что сейчас занятие, потом посмотрю, что там. Угу. А? Ну это вообще-то странно, вообще-то. Ну ладно. Вы настройки системы сделали, те, которые вот я говорил там в системе Windows? Да, делала. Вот эти настройки делали, да? Да. Ну вот. Ну вот.

* **Анализ результатов верификации**

Значит, мы сейчас что сделали? Мы сейчас задали текущей моделью для решения задачи идентификации и решили эту задачу. Но результаты ещё не смотрели. В режиме 4.1.2. Смотрим на результаты. Это делается подсистеме 4.1.3. Вывод результатов распознавания. К этому же относятся идентификация, я сказал, диагностика, прогнозирование. Тут есть довольно много разных форм, в которых можно эти результаты выводить. Но мы смотрим на самые простенькие. 4.1.3.1 форма. Вот смотрите, значит, слева у нас объект обучающей выборки указан. А справа классы, к которым этот объект отнесён системой, и уровень сходства с этими классами этого объекта. А здесь вот есть колоночка, где написано факт, и там стоит птичка. Обратите внимание. Если объект фактически относится, вот монитор один к классу монитор и к классу элемент компьютера, то там стоит птичка. Вам видно, а? Да, видно. Вот. Ну, прежде всего, давайте вот я вас спрашиваю, ребята, как вы думаете, почему на класс монитор этот объект монитор один больше похож, чем на класс элемент компьютера? Ответьте мне, пожалуйста. Вот сейчас мы в диалоге находимся. Лабораторное занятие, я задаю вам вопрос. Как вы считаете, почему, по какой причине монитор один больше похож на монитор на 90%, чем на элемент компьютера? Кто хочет получить зачёт автоматом, отвечайте, давайте. Не сидите молча. Ну я думаю, что монитор один, а комплектующих много. Абсолютно правильно. Потому что класс элемент компьютера, он создан, обобщённый этот образ, создан не только на основе мониторов, они там тоже присутствуют, но ещё и другие элементы компьютера: мышки, клавиатуры, тоже присутствуют в образе элемент компьютера. То есть он более размазанный такой, не так сильно похож на монитор, потому что он ещё похож на мышку, там, ещё похож на клавиатуру. В тех долях, которые они там вот присутствуют. Понимаете, да? Значит, вы совершенно правильно сказали. А почему монитор похож не на 100%, а на 90? Он же должен быть на 100% похож на монитор. По идее так. Потому что монитор - это тоже обобщённо. Мониторы тоже разные бывают. Абсолютно правильно. Вот, мы так неплохо всё понимаете. Смотрите, потому что они разные. Конечно. Вот. Ну и есть ещё и другая причина. Потому что на самом деле, когда создаётся обобщённый образ, то там, это не информационно-поисковая система. То есть она не будет похожа на 100%, потому что там ещё другие есть объекты, имеющие сходные признаки. И общая картина такая получается. А вот почему клавиатура, он похож на клавиатуру, монитор один оказался, на телефон, на телефон догадайтесь, почему он похож? Ну, наверное, потому что с телефона тоже можно вводить что-то. Ну, вообще-то, когда вот я эту табличку студенты делали, они телефон отнесли к средствам связи, а не к компьютерам. Но мы-то понимаем с вами, что это компьютер для связи. Правильно? А также на нём можно и другие функции выполнять, какие на компьютерах выполняются. Можно играть там, можно что-то там посмотреть PDF-файл, можно даже посчитать что-то. Правильно? То есть получается, что это вообще-то, на самом деле, компьютер, конечно, телефон. И у него есть, естественно, свой дисплейчик. Что очень сближает. Ну, то есть система просекает, что мониторы и телефоны похожи. Но тем, что у них есть дисплей. Помните, да, дисплей? Признаках. Вот они. Признаки эти. И по этим признакам они оказались достаточно сходные. Вот. То есть это, конечно, ошибка, что формально, формально это ошибка. Мышка один - это не телефон. Она не относится к этому классу. Но ошибка это с уровнем ниже 40% уровня сходства. Мы там видели вот в этом графике, что выше 40% только истинные решения. А от 30 до там до 40 и истинные, и ложные, да? Помните? Вот вам, пожалуйста. Вот это ложное решение. Вот оказалось, что мышка похожа на монитор. Ну, наверное, она похожа тем, что она пластмассовая, провода, кнопки там есть. Вот, наверное, этим похожа. Но она мало похожа, около 10%. То есть ложноположительное решение с низким уровнем сходства. Вот мы эту кривульку вот, это решение дало вклад вот в этой области. Вот. Это я, кстати, не знаю, почему происходит. По-моему, это из-за Вебокса. Такое подозрение. Потому что я когда вот просто запускаешь эти режимы все, то такого не происходит эффекта. Ну зависло. Ну подумаем. Вот. Короче говоря, вы поняли уже, что у нас есть критерии, ребята, когда мы на эту кривульку смотрим, на эти графики, то у нас появляется критерий, как интерпретировать результаты идентификации. Если у нас уровень сходства ниже 30%, модели инф3 именно, то тогда это ложное решение, неправильное. Если выше 40, тогда точно правильное. Если от 30 до 40, то может быть и правильное, и неправильное. Если больше 35, скорее что правильное, если меньше 35 до 30, скорее что неправильное. Вот. То есть мы видим, что у нас есть три диапазона уровня сходства положительного. Один диапазон ниже 30%, там только ложные решения. Другой диапазон выше 40%, там только истинные решения. И есть диапазон, где есть и истинные, и ложные решения, где-то там от 28 примерно до 40. Вот. В этом диапазоне есть и истинные, и ложные решения, причём при увеличении уровня сходства доля истинных решений растёт. Начиная с 35%, доля истинных решений больше, чем ложных. Вот. А до 30%, до 35 есть и истинные, и ложные, но доля ложных больше. Смотрим, значит, ну здесь ясно, что мы для того, чтобы сделать какие-то выводы, мы вычитаем из одной кривульки другую кривульку.

* **Идентификация как разложение в ряд**

Вот. То есть это, конечно, ошибка, что формально, формально это ошибка. Мышка один - это не телефон. Она не относится к этому классу. Но ошибка это с уровнем ниже 40% уровня сходства. Мы там видели вот в этом графике, что выше 40% только истинные решения. А от 30 до там до 40 и истинные, и ложные, да? Помните? Вот вам, пожалуйста. Вот это ложное решение. Вот оказалось, что мышка похожа на монитор. Ну, наверное, она похожа тем, что она пластмассовая, провода, кнопки там есть. Вот, наверное, этим похожа. Но она мало похожа, около 10%. То есть ложноположительное решение с низким уровнем сходства. Вот мы эту кривульку вот, это решение дало вклад вот в этой области. Вот. Это я, кстати, не знаю, почему происходит. По-моему, это из-за Вебокса. Такое подозрение. Потому что я когда вот просто запускаешь эти режимы все, то такого не происходит эффекта. Ну зависло. Ну подумаем. Вот. Короче говоря, вы поняли уже, что у нас есть критерии, ребята, когда мы на эту кривульку смотрим, на эти графики, то у нас появляется критерий, как интерпретировать результаты идентификации. Если у нас уровень сходства ниже 30%, модели инф3 именно, то тогда это ложное решение, неправильное. Если выше 40, тогда точно правильное. Если от 30 до 40, то может быть и правильное, и неправильное. Если больше 35, скорее что правильное, если меньше 35 до 30, скорее что неправильное.

* **Анализ факторов (SWOT-анализ)**

Вот. То есть мы действуем согласно логике и схеме преобразования данных в информацию, её в знания. И смотрим результат решения задачи принятия решений, управления, исследования предметной области. Сейчас вот давайте дальше двинемся. Значит, прежде всего про прогнозирование давайте в двух словах скажу. Значит, когда мы смотрим вот на эту таблицу исходных данных, на базу исходных данных, то у нас сейчас статическая интерпретация. Я вам рассказывал, что есть две интерпретации: статическая и динамическая. И ещё есть универсальная терминология. Вот эту форму я показывал вам хелпы программного интерфейса. Вам объяснял, что шкалы могут быть текстовые и числовые. Текстовые могут быть номинальные и порядковые. Что текстовые сортируются с признаком уникальные значения из них вытаскиваются. Числовые находятся минимальное, максимальное значение и делится на заданное число диапазонов. Интерпретация такая: это описательные шкалы и градации, это классификационные. Описательные описывают объект с его свойствами, а классификационные - принадлежность к обобщающим категориям. Это, значит, у нас статическая интерпретация. Динамическая интерпретация: описательные шкалы описывают факторы, действующие на объект управления, а классификационные шкалы описывают будущие состояния объекта управления, в которые он переходит под действием этих факторов. Это динамическая интерпретация. И универсально я просто говорю, что это описательные шкалы и градации, это классификационные. Это такая терминология, которая не поймёшь там, динамическая интерпретация или статическая. Вот. Так вот, ребята, значит, раз у нас задача идентификации решается, то точно так же математически решается задача прогнозирования. Почему? А потому что ничем они друг от друга не отличаются, кроме способа интерпретации э свойств объекта и значений или факторов действующих и будущих состояний. Кстати, вот эта папочка, которая была создана, смотрите, а вот папка приложения, вот, здесь все базы, папка приложения. И в этой папке приложения создана папочка, и там графическая форма. Ребята, все формы записываются графические в папочке со своими именами, всё сохраняется. Но хотя можно, конечно, Alt Print Screen сделать и поместить её там в документ. Но всё есть. Если на компьютере экран, разрешение монитора ниже, чем 1920 на 1080, то система это определяет и тогда на экране показывает масштабированную форму. Ну, допустим, 1360 точек там или 1000, сколько там, 1000, 1000, ну да, 1333, я не помню сейчас точно. Вот. И тогда графическая форма масштабируется до размеров фактических размеров экрана и показывается масштабированном виде. Ну как вот примерно с помощью просмотрщиков изображений. А вот сюда записывается в оригинальном виде с полным разрешением. И здесь ещё с этим же именем будет форма скалет, то есть шкалированная, масштабированная. Вот. Поэтому, значит, если мы интерпретируем свойства объектов как значение факторов вот эти колоночки, а эти как результат действия этих факторов, тогда это будет динамическая интерпретация. В этом случае решение задачи идентификации уже можно рассматривать не как решение задачи идентификации, а как решение задачи прогнозирования. То есть задача идентификации и задача прогнозирования - это одна и та же задача математически, но разный способ интерпретации причин и следствий. Это я хотел сказать.

* **Кластеризация классов**

Теперь смотрим, как вообще решается задача идентификации. Какой алгоритм её решения? Каждый объект обучающей выборки описан кодами. Кодами классов, к которым он относится, и кодами признаков, которые у него есть. А коды признаков - это градации описательных шкал вот этих. Вот они эти коды, вот они. 37 там, вот они эти коды. 20, 24. А классов, соответственно, классификационных шкал коды. В моделях посчитано, какое количество информации содержится в каждом признаке о принадлежности объекта с этим признаком к каждому из классов. Если признак встречается у какого-то, какой-то признак встречается у объектов определённой категории чаще, чем в среднем, то его обнаружение несёт положительное, определённое количество положительной информации о том, что он относится к данному классу. Если же данный признак встречается в данном классе, но реже, чем в среднем, вот, то тогда обнаружение этого признака несёт информацию, что он скорее этот объект не относится к этому классу. Ну, например, к примеру, вот длинные волосы несут довольно много информации о том, что это не парень, а много информации о том, что это девушка. А вот брюки, они несут очень мало информации о том, что это парень, и очень мало информации о том, что это не девушка. Почему? А потому что они и так почти у всех, понимаете? Вот. То есть мы получаем положительное количество, отрицательное количество информации в признаке о принадлежности к каждому из классов, содержащееся в некотором признаке. А если объект, ну, принадлежность объекта к какому-то классу, если у объекта есть этот признак. А теперь представьте себе, что нам известно, что у объекта не один признак, а много. Ну тогда логичным является считать, что объект принадлежит к тому классу, о принадлежности к которому в его системе признаков, уже во совокупности всех его признаков, максимальное количество информации. Это уже мы подходим к понятию интегрального критерия. В системе Эйдос есть два интегральных критерия. Один называется сумма знаний, такой более простецкий. Это просто суммарное количество информации в признаках о принадлежности к данному классу. Значит, представьте себе, что мы берём какой-то объект, у него есть определённые признаки. И в этой вот матрице модели, я вам только что её показывал, у нас эта содержится информация о том, сколько там информации в каждом признаке о принадлежности к каждому классу. Мы их суммируем. То есть у нас колоночка вот такая матрицы модели, колоночка. Можно мы её обозначаем И, вектор И. И житого класса. А И - это у него координата. А L(i) - это вектор, описывающий объект. Как он его описывает? Он очень просто описывает его. Если признак, итый признак есть у объекта, то там в этом векторе единичка. А если нет этого признака, тогда там нолик. Что получается в результате? В результате, что когда мы хотим просуммировать количество информации в тех признаках объекта, которые у него есть, о принадлежности к житому классу, то мы должны, вот по этой формуле посчитать. То есть если там признак есть, то на единичку умножается количество информации из матрицы информативности в этом признаке о житом классе. А если признака нет, то на нолик. То есть фактически мы ничего не делаем. Можно было бы формулу не писать, а сделать это логикой. То есть если признак есть, суммировать количество информации, если нет, то не суммировать. Но тогда бы не было формулы. А формула чем интересна? Тем, что, когда вы будете изучать линейную алгебру, аналитическую геометрию, то вам должны будут рассказать, что это вот скалярное произведение двух векторов. Это угол между двумя векторами. Это скалярное произведение, оно рассчитывается совершенно одинаково, независимо от того, в каком пространстве эти вектора, является ли это пространство ортонормированным и даже является ли оно там, допустим, плоским или искривлённым. То есть это очень фундаментальное свойство пространства, то есть не пространства, а вот этих векторов, не зависящее от свойств пространства. Это очень важно, потому что когда пространство не ортонормировано, а оно всегда не ортонормировано, оно бывает иногда близко к ортонормированному. То есть оси бывают близко к перпендикулярному, но никогда точно не будет перпендикулярно пространство модели. И можно использовать расстояние Евклида, ну, вариант теоремы Пифагора для определения расстояния. Но на самом деле, если мы посчитаем взаимосвязи между классами, между признаками, то оказывается, что они взаимосвязаны. И реально никогда пространство не бывает ортонормированным. Поэтому применение такой меры расстояния, как евклидово расстояние, она является некорректным. А вот это скалярное произведение не связано с этим предположением, называется информационное расстояние, я его называл или межвекторное расстояние - это мера корректная. Кроме того, есть определение шума. Белый шум определяется как такой сигнал, у которого каждый элемент этого сигнала имеет корреляцию с любым другим элементом этого сигнала, равную нулю. Ну, вернее так, там говорится не так, что она равна нулю прямо, а вот она стремится к нулю при увеличении размеров этих отрезков этих сигналов, фрагментов этих сигналов. Вот. То есть при увеличении этих фрагментов, э корреляция стремится между ними к нулю. Что это значит? Это значит, что если у нас в исходных данных есть шум, и мы используем именно вот такой белый шум, который так вот определяется, то и мы используем эту меру расстояния, то мы практически, фактически мы подавляем этот шум, не учитываем этот шум при принятии решений. Ну, правда, при больших данных. То есть если будет много признаков у объектов, вот, то тогда можно сказать, что шум не будет особенно сказываться на решении задач. Потому что этот, вот эта мера расстояния, интегральный критерий, она является определением белого шума. Шум будет подавлен таким образом, если он именно вот такой, белый.

Теперь ещё такой момент интересный. Когда мы определяем понятие белого шума, то мы это делаем следующим образом. Мы говорим так: шум - это такой сигнал, корреляция любого фрагмента, у которого корреляция любого фрагмента этого сигнала с любым другим фрагментом этого сигнала стремится к нулю при увеличении размеров этих фрагментов. Ну то есть, грубо говоря, последующая часть сигнала не зависит от предыдущей. Нет закономерности там. Это и есть шум. Вот. Так вот, эта мера, этот интегральный критерий, эта мера расстояния между объектом и классом, она устойчива к шуму, потому что если там будут элементы шума и в самом векторе класса, и в самом объекте, то суммарно они дадут нулевой вклад. То есть получается, что как бы не играет роли, что эти сигналы зашумлены.

Вот. Так вот, ещё интересный момент есть такой в этом, что, ортонормированная мера, мера Пифагора применима только в ортонормированных пространствах, а эта мера применима во всех пространствах. Вот. То есть она гораздо более разумная, эффективная. Называется межвекторное расстояние или информационное расстояние, я называл его. Теперь, ещё такой момент интересный. Когда мы определяем понятие белого шума, то мы это делаем следующим образом. Мы говорим так: шум - это такой сигнал, корреляция любого фрагмента, у которого корреляция любого фрагмента этого сигнала с любым другим фрагментом этого сигнала стремится к нулю при увеличении размеров этих фрагментов. Ну то есть, грубо говоря, последующая часть сигнала не зависит от предыдущей. Нет закономерности там. Это и есть шум. Вот. Так вот, эта мера, этот интегральный критерий, эта мера расстояния между объектом и классом, она устойчива к шуму, потому что если там будут элементы шума и в самом векторе класса, и в самом объекте, то суммарно они дадут нулевой вклад. То есть получается, что как бы не играет роли, что эти сигналы зашумлены. Ну, правда, при больших данных. То есть если будет много признаков у объектов, вот, то тогда можно сказать, что шум не будет особенно сказываться на решении задач. Потому что этот, вот эта мера расстояния, интегральный критерий, она является определением белого шума. Шум будет подавлен таким образом, если он именно вот такой, белый.

Теперь ещё такой момент интересный. Когда мы определяем понятие белого шума, то мы это делаем следующим образом. Мы говорим так: шум - это такой сигнал, корреляция любого фрагмента, у которого корреляция любого фрагмента этого сигнала с любым другим фрагментом этого сигнала стремится к нулю при увеличении размеров этих фрагментов. Ну то есть, грубо говоря, последующая часть сигнала не зависит от предыдущей. Нет закономерности там. Это и есть шум. Вот. Так вот, эта мера, этот интегральный критерий, эта мера расстояния между объектом и классом, она устойчива к шуму, потому что если там будут элементы шума и в самом векторе класса, и в самом объекте, то суммарно они дадут нулевой вклад. То есть получается, что как бы не играет роли, что эти сигналы зашумлены. Ну, правда, при больших данных. То есть если будет много признаков у объектов, вот, то тогда можно сказать, что шум не будет особенно сказываться на решении задач. Потому что этот, вот эта мера расстояния, интегральный критерий, она является определением белого шума. Шум будет подавлен таким образом, если он именно вот такой, белый.

Теперь, ещё такой момент интересный. Когда мы определяем понятие белого шума, то мы это делаем следующим образом. Мы говорим так: шум - это такой сигнал, корреляция любого фрагмента, у которого корреляция любого фрагмента этого сигнала с любым другим фрагментом этого сигнала стремится к нулю при увеличении размеров этих фрагментов. Ну то есть, грубо говоря, последующая часть сигнала не зависит от предыдущей. Нет закономерности там. Это и есть шум. Вот. Так вот, эта мера, этот интегральный критерий, эта мера расстояния между объектом и классом, она устойчива к шуму, потому что если там будут элементы шума и в самом векторе класса, и в самом объекте, то суммарно они дадут нулевой вклад. То есть получается, что как бы не играет роли, что эти сигналы зашумлены. Ну, правда, при больших данных. То есть если будет много признаков у объектов, вот, то тогда можно сказать, что шум не будет особенно сказываться на решении задач. Потому что этот, вот эта мера расстояния, интегральный критерий, она является определением белого шума. Шум будет подавлен таким образом, если он именно вот такой, белый.

Теперь, ещё такой момент интересный. Когда мы определяем понятие белого шума, то мы это делаем следующим образом. Мы говорим так: шум - это такой сигнал, корреляция любого фрагмента, у которого корреляция любого фрагмента этого сигнала с любым другим фрагментом этого сигнала стремится к нулю при увеличении размеров этих фрагментов. Ну то есть, грубо говоря, последующая часть сигнала не зависит от предыдущей. Нет закономерности там. Это и есть шум. Вот. Так вот, эта мера, этот интегральный критерий, эта мера расстояния между объектом и классом, она устойчива к шуму, потому что если там будут элементы шума и в самом векторе класса, и в самом объекте, то суммарно они дадут нулевой вклад. То есть получается, что как бы не играет роли, что эти сигналы зашумлены. Ну, правда, при больших данных. То есть если будет много признаков у объектов, вот, то тогда можно сказать, что шум не будет особенно сказываться на решении задач. Потому что этот, вот эта мера расстояния, интегральный критерий, она является определением белого шума. Шум будет подавлен таким образом, если он именно вот такой, белый.

Вот. И вот мы видим, что разные объекты в разной степени похожи на обобщённый образ элемент компьютера, к которым относятся. То есть вот эти все объекты, где птичка, они относятся к этому образу. Но они похожи на него в разной степени. Вот это называется типичные объекты, а это называется исключения. Вот клавиатура один - это исключение. Она относится к этому классу, но она на него меньше всего похожа. А вот мышка один - это типичный объект. Она больше всего похожа на этот класс. Вот. И есть ложные решения. Пинг-понговый мячик похож на элемент компьютера. Но очень мало, видите? 5%. А вот телефон два похож на элемент компьютера на 28%. Ну это уже существенно. То есть это уже такая ошибка, которую надо учитывать.

Вот. То есть эта форма, она показывает нам результаты решения задачи идентификации. А если мы решаем задачу прогнозирования, то это будет результат решения задачи прогнозирования. То есть мы видим, что если у нас будут вот такие значения факторов, как у мышки один, то наиболее вероятным будущим состоянием будет состояние мышка, а также состояние элемент компьютера. А вот состояние телефон, средство связи, клавиатура, монитор, они тоже возможны, но с гораздо меньшей вероятностью. А вот состояние мяч, сумка, аксессуар, мебель, стул, вешалка, спортинвентарь, они вообще невозможны, потому что уровень сходства отрицательный. То есть это означает, что объект не перейдёт в это состояние. Вот так вот.

Есть обратная операция. Обратное преобразование Фурье, например. А вот форма 4.1.3.2, она тоже в системе Эйдос реализуется. Здесь слева видим не объекты распознаваемые, а прямо сами классы. И мы видим, что у нас по каждому классу справа объекты есть, которые больше всего похожи на этот класс. И по некоторым классам у нас есть один объект, например, в качестве примера. По некоторым два, по некоторым там, а вот по классам обобщающим у нас много объектов есть. Вот обобщающие классы. Видите? Элемент компьютера - это обобщающий класс. Много разных есть элементов: мышь 1, монитор 2, мышь 3, монитор 1. И они в разной степени похожи на этот обобщающий класс. Меньше всего похож на обобщающий класс элемент компьютера клавиатура 1, 57%. А больше всего мышка 1. Вот. А вот на средство связи похожи все телефоны. И почти что одинаково, 91-92%. А на спортинвентарь больше всего баскетбольный мяч похож, и футбольный, и теннисный. И меньше всего пинг-понговый мячик похож на спортинвентарь.

О чём это говорит, ребята? Какой вывод из этого можно сделать? Такой, что образ класса отражает статистику объектов, которые к этому классу относятся, их обобщённые свойства этих объектов. И всегда есть и объекты типичные для класса, и всегда есть исключения. Вот. Ну, допустим, берём мы создаём образ студента и образ студентки на основе студентов вот ваших групп. И есть вероятность, что у одного студента будут длинные волосы. Ну есть такой прикид там определённых подмножеств культурологических групп: музыканты, художники - это как бы у них профессионально. И вот, значит, что мы можем сказать? Что этот студент, когда мы посмотрим здесь образ студента, обобщённый образ студента, то он получится похожим на студента, ребят, он получится похожим на студента, не на студентку. Но всё равно будет уровень сходства с классом студент минимальный из всех студентов у него будет. То есть эти длинные волосы, они несут большой объём информации о том, что это не студент, а студентка. У него есть и другие признаки есть, которые больше суммарный объём информации, чем эти длинные волосы. То есть он всё равно будет похож на студента, а не на студентку. Но гораздо в меньшей степени будет похож, чем остальные студенты. Вот. И борода, наличие бороды - это признак очень несущий очень большой объём информации о том, что это студент. Вот. А вы помните, было Евровидение, где там была Кончита Вурст? Первое место заняла. Такая не поймёшь, что она такое. Ну, вроде как девушка, но с бородой. Такой коротенькой, но густой. Помните, нет, Кончиту Вурст? Да, была такая. Ужас какой-то. Ну, в общем, короче говоря, так. Бывает, конечно, и девушки с бородой, но но редко, понимаете, слава Богу. То есть этот признак, он у девушек встречается, но очень редко, понимаете? То есть он будет нести большой объём информации о том, что это парень, и большой объём информации о том, что это не девушка. Ну так же, как вот длинные волосы у парня. То есть, соответственно, Кончита Вурст, она будет на девушек не очень похожа получится. Хотя по остальным признакам она, возможно, и похожа на девушку. Я там не разглядывал особенно, мне прямо что-то дурно стало, когда я её увидел. Вот. Но я не исключаю, что она по остальным признакам похожа, но этот признак он просто вот, я не знаю, делает её очень похожей не на девушку, а на парня. И она будет на девушек похожа в минимальном, с минимальным уровнем сходства из всех девушек. Вот. Ну вот так вот.

Это я всё вам рассказал про решение задачи идентификации, про достоверность модели, решение задачи идентификации.

**7. Решение Задач с Использованием Модели (продолжение)**

* **Принятие решений (Обратная задача)**

Ну теперь давайте другие задачи решать. Вот. То есть мы изучили, как осуществляется формализация предметной области, синтез и верификация моделей. Как осуществляется, по каким критериям оценка их достоверности, как решается задача идентификации и прогнозирования. Прогнозирование ещё не рассказывал я. И как решается задача принятия решений, управления и исследования предметной области. Сейчас вот давайте дальше двинемся. Значит, прежде всего про прогнозирование давайте в двух словах скажу. Значит, когда мы смотрим вот на эту таблицу исходных данных, на базу исходных данных, то у нас сейчас статическая интерпретация. Я вам рассказывал, что есть две интерпретации: статическая и динамическая. И ещё есть универсальная терминология. Вот эту форму я показывал вам хелпы программного интерфейса. Вам объяснял, что шкалы могут быть текстовые и числовые. Текстовые могут быть номинальные и порядковые. Что текстовые сортируются с признаком уникальные значения из них вытаскиваются. Числовые находятся минимальное, максимальное значение и делится на заданное число диапазонов. Интерпретация такая: это описательные шкалы и градации, это классификационные. Описательные описывают объект с его свойствами, а классификационные - принадлежность к обобщающим категориям. Это, значит, у нас статическая интерпретация. Динамическая интерпретация: описательные шкалы описывают факторы, действующие на объект управления, а классификационные шкалы описывают будущие состояния объекта управления, в которые он переходит под действием этих факторов. Это динамическая интерпретация. И универсально я просто говорю, что это описательные шкалы и градации, это классификационные. Это такая терминология, которая не поймёшь там, динамическая интерпретация или статическая. Вот. Так вот, ребята, значит, раз у нас задача идентификации решается, то точно так же математически решается задача прогнозирования. Почему? А потому что ничем они друг от друга не отличаются, кроме способа интерпретации э свойств объекта и значений или факторов действующих и будущих состояний. Это я хотел сказать.

Теперь давайте, есть, конечно, различная терминология в различных областях. Вот в одной области задача идентификации называется идентификация, в другой - распознавание, в третьей - диагностика, в четвёртой - прогнозирование. То есть это называется по-разному. Сама задача идентификации называется по-разному: распознавание, идентификация, классификация, диагностика, прогнозирование. Здесь я не написал варианты. Вот. Смотрим теперь решение задачи принятия решений. Предположим, что у нас, я так надеюсь, у вас фантазия развита, ребят. Представим себе, что эта задача у нас задача прогнозирования решалась. То есть это не признаки объектов, а значения факторов. И мы хотим решить обратную задачу, э обратную задачу прогнозирования. Почему я говорю, что принятие решения является обратной задачей прогнозирования, ребят? Потому что... Вот опять же написано, где будет создана папочка для выходных форм графических. Вот, допустим, берём мы элемент компьютера, модель инф3, то, что я вам говорил. Видим, вот такие вот значения факторов обуславливают переход объекта моделирования, объекта управления в состояние, соответствующее вот этому классу. А эти препятствуют. То есть это классический SWOT-анализ. Ребята из КГУ, вы изучали SWOT-анализ? У вас было это? Вы ж там экономисты, должны были изучать. Скажите, изучали вы, нет? Ну вообще... Ребят, вы изучали SWOT-анализ? Изучали? Молодцы. Да или нет, я не понял только. Изучали. Короче говоря, значит, обычно строится SWOT-матрица, где слева располагаются все факторы, способствующие достижению некоторого результата, справа - препятствующие. Если спросить э дипломников, э магистрантов, которые защищают диссертацию, э соискателей учёных степеней, которые показывают на слайдах SWOT-матрицы, спросить их: "А где вы взяли вот эти вот коэффициенты нагрузки? Вот насколько этот фактор способствует этому переходу объекта вот в это состояние? И насколько этот фактор препятствует переходу в это состояние? Откуда вы взяли вот эти числа?" Обычно вообще там чисел нет, там просто перечислены факторы. Но иногда бывает там что-то вроде баллов там, пять баллов, там, три балла. Да? Ответ одинаковый всегда, ребята: экспертным путём. А кто был экспертом? Да, если так между нами, то сами же и были экспертами. Правильно? Вот. Ну, может, там кто-нибудь и был экспертом. Вот. Система Эйдос является единственной на данный момент системой, которую вообще можно найти в интернете, которая обеспечивает автоматизированный SWOT-анализ. То есть она сама вычисляет количество информации, содержащееся в том или ином значении фактора о переходе объекта под действием этого фактора в определённое будущее состояние. И можно это графически тоже представить себе. И туда это вот всё записывается. Смотрите. SWOT-диаграмма классов модели инф3, приложения. Ну это стандартное учебное приложение, лабораторная работа 3.03, локальная, которая прямо в инсталляции есть. Шкала обобщающий класс, сам класс, градация шкалы, элемент компьютера. Наиболее характерным для этого класса является наличие проводов, наличие кнопок, пластмасса, материал, форма округлая, квадратная, средний размер, цвет чёрный. А наиболее нехарактерными является, что наличие проводов нету их, кнопок нету, форма круглая. Вот, размер один. Смотрите, шкала. Большого размера - это не элемент компьютера, а среднего размера - это элемент компьютера. Ножки есть у него. Если ножки есть, то это не характерно. Сложная форма не характерна. Вот. Вот так мы можем про каждый класс узнать.

Значит, почему задача принятия решения является обратной по отношению к задаче прогнозирования? Потому что при прогнозировании мы по значениям факторов определяем, в какому будущему состоянию, какое будущее состояние перейдёт объект моделирования. А при принятии решения мы наоборот, по заданному целевому состоянию определяем, какая необходима система значений факторов, чтобы объект перешёл в это целевое состояние. То есть это чистая классическая обратная задача. Все ли модели интеллектуальных систем обеспечивают решение этой обратной задачи? Я вам скажу, далеко не все. Далеко не все. Почему? Значит, есть одна особенность этих моделей, которая должна быть обязательно, чтобы эта задача в принципе могла решаться. Может быть, система её не решает, но модель позволяет. Просто не сделали режим. А может быть, модель вообще в принципе не позволяет её решить. Значит, вот есть такой метод распознавания - метод К ближайших соседей, где обобщённые образы классов вообще не создаются. А просто результатом распознавания является сравнение конкретного объекта с другими конкретными объектами обучающей выборки. Вот, допустим, берём вашу группу, распознаём, кто у нас вот этот вот конкретный студент? Кто вот он у нас? Результат распознавания. Вот. Вот я сейчас беру и смотрю. Здесь это начинает немножко... Так. Вот Марина есть, да, староста, да, Подгорная? Вот я смотрю и думаю, э, какой результат распознавания Марины по её имиджу? И вижу, что она похожа на Екатерину, на Кристину, на Ольгу, на Викторию. А вот на Олега там, на Даниила, э, вот, на Артёма не похожа. Это называется метод К ближайших соседей. То есть просто перечисляется какое-то число объектов обучающей выборки, на которые она больше всего похожа и больше всего не похожа. Или или даже только те, на кого она похожа, перечисляются. Это вот и есть результат распознавания. А как мы поступаем, когда модель позволяет создавать классы? Создаются обобщённые образы классов на основе ряда примеров. И объект сравнивается не с другими объектами распознаваемыми, а он сравнивается с обобщёнными образами классов. Так вот, если обобщённые образы классов есть, как вот вы видели в этих формах экранных, которые я показывал, то тогда можно решить обратную задачу. То есть можно взять э модель... Сейчас я вам покажу, как это делается. Вот берём модель какую-то, в которой мы решаем задачу. Вот, допустим, модель инф3. Вот нас интересует, какие значения факторов необходимо использовать, чтобы объект перешёл в состояние, соответствующее вот этому классу четырнадцатому, обобщающий класс, элемент компьютера. Что для этого нужно сделать, ребят? Для этого мы просто сортируем вот эту базу данных, соответствующую этой модели. Ну это таблица просто, база данных, матрица. Сортируем по этой колоночке вот, соответствующей этому классу, в порядке убывания. Ну, естественно, все целиком строчки, целиком записи сортируются. Это можно делать не физически, а логически, просто индексный массив формируется. Это происходит гораздо быстрее, чем физическая сортировка, особенно когда база большие. И вот э результат этой сортировки у нас такой, что больше всего способствует переходу объекта в состояние, соответствующее этому классу, вот такое значение фактора. Поменьше такое. А в самом конце будут те, которые просто препятствуют. В самом в самом конце будут те, которые препятствуют в наибольшей степени, потом те, которые препятствуют несколько в меньшей степени. Понятно, да, ребят? Как вы оцениваете? Ну, в общем, понятно. Так вот я могу вам сказать, что эта задача решается мгновенно. Вот всё, она решена уже. И даже на больших моделях она решается мгновенно. Почему, знаете? Ну, индексный массив создаётся оптимальным методом двоичного поиска. Ну оптимальные методы математические используются. И получается это очень быстро, очень хорошо. То есть задача принятия решения решается мгновенно фактически. Вот. Если же мы будем пытаться её решать путём прогнозирования, то есть будем многократно осуществлять прогнозирование при разных значениях факторов, при разных сочетаниях значений всех факторов. У нас будет L в степени n по М число сочетаний значений факторов. У нас могут быть там сотни тысяч раз нужно спрогнозировать, чтобы узнать, что будет завтра. А мы можем это сделать такое число прогнозирования, только если неделю потратим. То есть мы не сможем за необходимое время спрогнозировать на необходимый период. То есть нужно большее время. А вот таким образом, как решение обратной задачи, это получается быстро очень.

* **Исследование предметной области**

Смотрим теперь следующее, что можно ещё узнать интересного. Мы можем узнать, как влияет любое значение фактора на результат. Опять нет папочки. Это называется инвертированные SWOT-диаграммы. Такого в науке не было, это я предложил инвертированные SWOT-диаграммы. Ну, это сходно, знаете, с чем? Семантический потенциал фактора. То есть это что значит? Какой смысл этого вот признака? Вот признак, допустим, материал пластмасса. О чём этот признак говорит? О том, что скорее всего, это элемент компьютера, возможно, мышка, телефон, средство связи, клавиатура или монитор. И ни в коем случае это не мебель, не аксессуар, не сумка, не спортинвентарь, не мячи, не стулья, не вешалки. То есть это смысл этого признака. Ребята, система сама это вычисляет этот смысл на основе эмпирических данных. Как и SWOT-диаграмму. И знает это по всем признакам. Значит, фактически, когда эта задача решается, просто сортируется строка матрицы модели. Берётся какая-то модель, вот строка, допустим, там материал там или что там, пластмасса, да, вот, допустим, семнадцатая. Вот. Просто берётся эта строка и сортируется в порядке убывания количества информации в этом признаке о принадлежности классов к этим вот, то есть объектов к этим классам. И классы располагаются в порядке убывания количества информации о принадлежности к ним в этом значении фактора. Понятно, да? То есть это делается всё просто, ребята, просто, понимаете? Что вот, собственно говоря, интересно тоже. Следующее. Мы можем посмотреть, а что у нас за элемент компьютера с точки зрения нейронных сетей? Мы можем увидеть э в интерпретации нелокального нейрона это. Нелокальные нейронные сети я предложил в 2003 году в статье. Это один слой нелокальной нейронной сети. Система Эйдос позволяет многослойные модели строить. Несложно, причём, это делается совершенно. Я это делал ещё там в девяностых годах. Так вот, ребята, смотрите. Значит, здесь у нас вверху классы, и они в этой интерпретации нейросетевой - это нейроны. А здесь у нас значения свойств или значения факторов. В этой интерпретации нейросетевой - это рецепторы и нагрузки на этих рецепторах. И мы что делаем? Мы вычисляем все связи каждого рецептора со каждым из нейронов, все, абсолютно все, какие только есть. А потом сортируем в порядке убывания модуля этих связей. И рисуем те, которые помещаются на экране при данных параметрах. Здесь вот у нас эти параметры перечислены. Значит, я могу эти параметры поменять. Вот я, допустим, могу сказать, что вот у нас нейронов 16, а всего их 14, хватит. Максимальное число связей - 1000. Диапазон отображаемых рецепторов - 50. Максимальное число отображаемых рецепторов - 16. Вот. Я сейчас возьму и задам 50. Вот. И посмотрим, что получится. Получится, что там ничего не поймёшь. То есть уже наползают наименования, но коды ещё видны. И я могу вам сказать, что оно так, в общем-то, симпатично выглядит, но на самом деле уже нечитабельно и нет смысла особого в этом, потому что вот эти вот связи, которые здесь справа - это связи, которые практически ни на что не влияют. То есть это можно такое вот для того, чтобы пугать детей показывать такие диаграммы. Вот здесь видите, 100% отображено информации, которая есть в сети на данный момент. А если здесь мы зададим стандартное это значение, когда хорошо всё прорисовывается, там получится не 100%, а 60%. А бывают задачи, когда вот рисуешь такую вот подмножество, а там получается 3% там или 5%. Вот. Смотрим дальше. Интегральные когнитивные карты. Сразу я просто беру и... Нейронов беру пять, вот, а рецепторов беру семь. Потому что иначе мы не поймём, что там. Ага. Видите, значит, когда нарушаешь последовательность обработки, преобразования данных в информацию, её в знания, она нам высказывает всё, что она нас думает система. Она говорит: "Перед этим, чтобы вы хотите сейчас посмотреть, надо вот эти режимы выполнить". Ну давай их и выполним. Значит, э режим 4.2 2.1 - это расчёт матрицы сходства в разных моделях. Посчитали. Теперь мы можем посмотреть эту матрицу сходства в виде базы данных, она открывается в Экселе. Вот. Можем посмотреть в виде когнитивной диаграммы. Тоже куда она будет записана. Здесь мы видим уже решение задачи, ребята, другой. Не сравнение конкретного объекта с обобщёнными образами классов, а сравнение самих обобщённых образов классов друг с другом. И можем это же сравнение провести в форме агломеративной дендрограммы кластерного анализа провести.

**8. Организационные Моменты и Заключение**

Так, ребята, у нас сейчас, похоже, что конец занятия, да? 11:10. Значит, ребята, из КГУ, вы можете уже... Так, вот дендрограмма. Дендрограмма. Из КГУ, ребята, у вас конец занятия, поэтому вы можете, в принципе, покидать совещание, но можете посмотреть ещё несколько минут. 5 минут будет. Вот мы здесь видим, ребята, вот эта шкала различий, а это перечислены классы. И мы видим, насколько какие классы друг на друга похожи и не похожи. И видим, что эти классы все образуют кластеры. Вот стул и мебель, допустим, кластер. И ещё туда стул и вешалка входят. Вот это кластер, видите, вот он образуется. Вот. И это образует большие кластеры, большие, на самом верхнем уровне различий, сходств и различий. С одной стороны всё, что связано с компьютерами, а также с телефоном, а на другой всё, что связано с мебелью и всякими сумками там и спортинвентарём. Вот так вот. Это называется конструкт. Про конструкт я вам, кажется, рассказывал на предыдущих занятиях. Всё, ребята, для КГУ конец занятия. Надеюсь, вам понравилось. И сейчас, собственно, и в аграрном тоже конец занятия. Значит, пожалуйста, если какие-то есть вопросы, можете спросить, ребята. Давайте. Какие есть вопросы? Пожалуйста, спрашивайте. Ну, у студентов всегда два вопроса. Один спрашивает, когда мы когда закончится занятие, а второй - где взять там денег, я не знаю, так что-нибудь такое, да? А ещё вопросы есть вот по теме самой занятия? Ну кто-нибудь отвечает, а то что вы все так сидите там? Вы это там вообще есть вот Марина Подгорная, староста? Вопросов не имеем, всё понятно. Спасибо большое. Пожалуйста. Значит, тогда мы в следующем занятии продолжим решение задач исследования моделируемой предметной области путём исследования её модели. Как вы догадались, мы уже решили задачу идентификации, прогнозирования, принятия решений. И сейчас уже занялись исследованием моделируемой предметной области путём исследования её модели. Вот эти формы, которые я последние вам показывал, они к этому именно относятся. Но я показал не все. Сейчас мы в следующем занятии мы продолжим эту тему. Именно вот исследование объекта путём исследования её модели, не объекта, а моделируемой предметной области путём исследования её модели. Всё, ребята, всего самого хорошего, до свидания. До свидания. До свидания. Спасибо, всего доброго. До свидания.