**ФГБОУ ВО «Кубанский государственный аграрный университет имени**

**И.Т. Трубилина»,**

**Российская Федерация**

**Колесников Роман Юрьевич, ПИ2102**

**roman563412@gmail.com**

**Семинар №2. Рассмотрение системно-когнитивного анализа и его программном инструментарии.**

**Заголовок**

Системно-когнитивный анализ с использованием системы Эйдос: Обзор, история, функционал и применение

**Резюме текста**

Семинар посвящен системно-когнитивному анализу (СК-анализу) и его программной реализации — системе Эйдос, разработанной под руководством профессора Луценко Е.В. СК-анализ представлен как общий метод и теория, включающая математические модели, алгоритмы и структуры данных, тогда как Эйдос является одним из его программных инструментариев.

Обсуждается история разработки системы Эйдос, начиная с ранних версий на языке Clipper для DOS-совместимых компьютеров, через различные реализации на других языках и платформах (включая Wang 2200C), до текущей версии на языке Alaska (объектно-ориентированный диалект Clipper/XBase++) под Windows. Подчеркивается, что выбор языка программирования для реализации может меняться со временем, и текущая версия использует Alaska 2.0+ и библиотеки Express++.

Затрагиваются текущие технические проблемы, связанные с переносом сайта и хостинга системы с Windows на Linux, что временно затрудняет обновление системы и доступ к облачным приложениям из-за нестандартного порта FTP (2211 вместо 21).

Демонстрируется основной принцип работы Эйдос: преобразование исходных данных в информацию (путем выявления причинно-следственных связей, согласно концепции Шенка-Абельсона) и далее в знания (при использовании для поддержки принятия решений и управления). Система поддерживает адаптивное управление (согласно принципу дуальности Фельдбаума), позволяя моделям обучаться и адаптироваться в процессе работы.

Рассматривается процесс формализации предметной области, включая создание классификационных и описательных шкал, кодирование данных и формирование обучающей выборки. Приводится пример анализа данных предприятий (финансовые показатели) для оценки их состояния (кризис/норма).

Обсуждаются различные типы данных и программные интерфейсы в Эйдос (табличные, текстовые, графические), возможности обработки текстов и графики (оцифровка по контурам, спектрам, пикселям). Упоминается пример приложения №130 и связанная с ним научная публикация.

Кратко описываются этапы анализа в системе: синтез моделей (статистических и СК-моделей знаний), их верификация (проверка достоверности, в том числе с использованием бутстреп-подхода), и решение задач (идентификация, прогнозирование, поддержка принятия решений, исследование предметной области). Упоминаются инструменты визуализации и интерпретации результатов, такие как когнитивные диаграммы, дендрограммы и конструкты.

Детальная расшифровка текста:

1. Введение и представление темы

Начало семинара

Да. Да-да, конечно. Здравствуйте, уважаемые коллеги. Сегодня 25 апреля 23-го года, 17 часов. У нас научный семинар под руководством Анны Владимировны Коваленко, зав кафедры искусственного интеллекта и обработки данных, доктора технических наук, доцента. И вот выступаю я, профессор Луценко Евгений Вениаминович.

Тема семинара: СК-анализ и система Эйдос

Это второй семинар по рассмотрению системно-когнитивного анализа и его программного инструментария, в качестве которого в настоящее время выступает система Эйдос.

2. Концепция СК-анализа и системы Эйдос

СК-анализ как общий метод

Почему я так витиевато выразился, что в настоящее время? Потому что системно-когнитивный анализ — это общий метод. Это теория, модели математические, алгоритмы, структуры данных. А программная реализация может быть, в общем-то, самая различная, на различных языках.

История языков программирования и Эйдос

И вот есть такая особенность, что когда начинаешь разрабатывать программную реализацию большой системы, развивающейся, то когда заканчиваешь эту реализацию, то никто не знает, что это за язык программирования, на котором она сделана.

Я так пошучу немножко. Ну вот сейчас вот какой самый популярный? Python, наверное, да? Вот сейчас если начать разработку на Python, то когда она будет закончена, никто не будет знать даже, что такое этот Python, вообще никто и не слышал там. Будут думать, что это змея там.

Ну как вот сейчас вот скажешь, что такое Марс? Ну все говорят: "Ну это такая шоколадка Марс". А то, что есть планета, мало кто вспоминает.

3. Технические моменты: Запись и демонстрация экрана

Включение записи

Так, значит, сейчас мы должны запись включить. Да? Вы включили запись.

Включение демонстрации экрана

И демонстрацию экрана. Что-то я подзабыл, где тут эта демонстрация экрана. Действия, да? Подскажите, где демонстрация экрана? Что-то я...

Вверху, вот справа кнопочка красная "Выйти", а чуть левее красной кнопочки...

А, поделиться. Всё, всё, всё, спасибо. Спасибо.

4. Уточнение по языку реализации Эйдос

Вопрос о языке

А тогда сразу вопрос. Вот вы сказали, что на каком языке написано? Так на каком языке написано-то?

Ответ: Clipper и история версий

На языке Clipper. Изначально была первая версия, которая на IBM-совместимых компьютерах, на языке Clipper была написана. Вот. А последующие версии, потом э-э были, ну, до этого были реализации не на языке Clipper, на других языках.

5. История разработки системы Эйдос (Детально)

Ранние этапы и платформы

Это коротко об этом говорится на вот этой страничке.

Вот. Сейчас попробую вспомнить, как здесь всё это делается.

Вот на этой страничке здесь есть краткая такая история, тезисно изложена. Этапы разработки системы Эйдос. Первый этап — это 79-92 годы. Это была ДОСовская реализация прямо вот на... А, нет, первая 79-92 — это были на различных языках программирования и на других компьютерах, не на IBM-совместимых. Ну, в частности, на компьютерах Wang 2200C.

ДОС-этап (Clipper)

Потом в последующем, второй этап — это я называю ДОСовский этап. Вот, 92-2012 годы.

Переход на Windows (Alaska)

Потом я понял, что... Да я это гораздо раньше понял, но пришлось немножко задержаться. И в 2012 году появилась Windows Vista, Windows 7, семёрка. Вот. И я разработал систему на новую версию на языке, который является языком Clipper на под Windows. Это язык Alaska 1.9. Были другие варианты. Я их рассматривал, пытался на Джаве, на Delphi for PHP. Вот. Ну потом, как бы, нельзя сказать, что не получалось, но потом как бы я понял, что не успею к учебному году, к сентябрю. И, в общем, имея большой объём наработок на языке Clipper с предыдущей версии, я взял... Там тоже были варианты. Был Harbour язык такой, который, по сути, C является, предкомпилятор C, использован язык XBase или dBase его ещё называют, древний язык. Это, э-э, вы, конечно, слышали. Ну вот FoxPro, например, это язык dBase. Clipper, Clarion. Вот. Ну это вот язык FoxPro, можно сказать, похожий очень. Только компилятор истинный. Вот. А потом, значит, когда возникла Windows, то он разделился на несколько разных версий. И вот одна из них Harbour — это компилятор C++ использовался и предкомпилятор, который язык dBase использовал. А потом прямо вот компилятор C++ классический. А потом, значит, и ещё некоторые фирмы появились, Alaska Software Corporation, например, которая разработала полностью совместимый язык с языком Clipper, назвали его Alaska, XBase++. Вот на нём я и разработал.

Современный этап (Alaska 2.0+)

Потом новая версия появилась языка этого, когда Windows 10 стала распространяться. И, значит, это вот современная версия системы Эйдос, написана на языке Alaska 2.0+ плюс библиотеки Express++. Вот. И, в общем, на этом языке она сейчас вот и существует.

Попытки на других языках (C#)

Я пробовал на других языках сейчас современных, э-э, на C# были попытки. Вот, но пока ничего такого, чтобы показывать кому-то нет. Там на уровне, ну, в общем, не завершённые попытки эти.

6. Этапы разработки и доступ к материалам

Обзор этапов

Вот. Таким образом, это вот этапы.

Ссылки на задания и приложения

Теперь давайте посмотрим, где есть у нас примеры. Это вот ссылка на задание для студентов, которое я упоминал, аспирантов, магистрантов, прочих учащихся. И здесь у нас есть ссылочка на уже имеющиеся облачные Эйдос-приложения, а также на э-э облачные э-э записи в облаке видеозанятий.

Ссылки на записи и методические материалы

Вот. Записи в облаке. Значит, я вот так вот сошлюсь на них. Вот. И на некоторые источники методической, учебно-методической информации тоже здесь сразу всё это прошлю сюда.

Ссылка на видеозанятие (обработка текстов и графики)

И здесь есть одно, одна ссылочка на занятие. Я вот даже вот так его пошлю. Это две пары, то есть сдвоенная пара, где рассматривается обработка текстов в системе Эйдос и обработка графики.

Типы данных и интерфейсы в Эйдос

Но в каждом типе данных в системе Эйдос, там есть шесть, в системе Эйдос есть шесть различных программных интерфейсов с различными типами внешних исходных данных: табличными, файловыми текстовыми и файловыми графическими. И каждый из них несколько вариантов имеет, потому что три типа данных, а интерфейсов шесть. И вот, скажем, графический интерфейс оцифровки изображений, он имеет три варианта: оцифровка по контурам, по спектрам и по всем пикселям. Я вам покажу по спектру, как наиболее зрелищный вариант, но сошлюсь на работы, которые все эти варианты демонстрируют.

7. Облачные приложения и текущие технические проблемы

Количество приложений и причины недоступности

Вот. Так вот, среди уже созданных приложений мы здесь вот видим сейчас на данный момент их 381 это приложение. На самом деле их гораздо больше. Вот, но некоторые не размещены в облаке по различным причинам, в том числе по техническим причинам.

Проблема с хостингом и портом FTP

Сейчас мой сайт переехал с хостинга Windows на Linux. Вот, но переехал пока не до конца, так немножко коряво. Заключается эта корявость в том, что порт, на котором прописан хостинг 2211 — это нестандартный порт, а стандартный 21. Вот. И поэтому у всех пользователей в мире, которых довольно много, вы видели, что система во всём мире используется, запускается. Значит, у них программа обновления была на стандартный порт. И они новую версию не могут получить. Я, значит, упросил нашего начальника центра информационных технологий, чтобы он перенёс этот хостинг Linux на порт 21, чтобы у них обновления прошли системы. Это одна причина, а вторая причина такая, что э-э не только обновление, но и скачивание облачных Эйдос-приложений вот этих вот, и запись облачных Эйдос-приложений тоже сейчас не работает. Хотя в языке программирования, конечно, есть этот возможность задавать порт конкретной работы с хостингом, значит, но с FTP сервером, но э-э вот по этим причинам пока это не работает. Версию я не меняю, именно ориентируясь на то, что будет этот порт задан. Но для того, чтобы его задать, сначала удалят э-э физически, уничтожат старый сайт. А новый он давно работает уже, но он работает вот в таком режиме, просто сайта, а не хранилища облачных приложений и сервера обновлений.

Последствия для пользователей

Так что вот такие причины. То есть сейчас я не буду показывать, поэтому скачивание из облака приложений. Это будет на следующей неделе. Сейчас этого нет специалиста на месте. Вот. Сделаем мы, вот.

8. Демонстрация приложения №130

Поиск и запуск примера

Сейчас посмотрим, что у нас тут. Вот, это у нас... Это не то. Не то. Хм. Вот это да.

Вот, смотрим приложение 130.

Ссылки на приложение и публикацию

Значит, вот ссылочка на сервере приложений. Ссылочка на нашу публикацию с Анной Владимировной и Али Хусеином Мухаммедом Али Хусеиновичем. Сейчас я ссылочку эту сюда отправлю. Эта статья издана в журнале довольно престижном, Вестник Пермского университета. И вот по этой статье и будет вот моё сейчас выступление.

Запуск локальной версии (из-за проблем с хостингом)

Значит, кратенькое. Значит, это у нас приложение 130. Захожу я сейчас на сайт свой, но и то он не на хостинге, а я уже сказал, предупредил, что будем смотреть его в оффлайн-варианте как бы, по тем причинам, которые я сказал.

Копирование данных примера

И вот сюда скопируем данные из этого приложения, 130-го приложения. Ну, в ближайшее время, я думаю, на следующей неделе сайт будет переедет до конца на этот хостинг, на новый. И тогда мы будем уже в самой системе Эйдос иметь возможность посмотреть.

Запуск системы и интерфейса

Значит, э-э то есть, запустив систему, сможем скачать это приложение, и сразу будет запущен соответствующий программный интерфейс. Информация о котором есть в самом, на самом хостинге о том, какой интерфейс использовался. Вот API TXT API 2322 использовался. Значит, это не для человека, а для системы информация. То есть система запустит соответствующий интерфейс при скачивании этого приложения, и оно будет установлено.

Ручная загрузка данных (обход)

Но сейчас мы это сделаем вручную. Так как это делается, значит, когда мы скачиваем просто исходные данные.

Структура исходных данных

Вот какие у нас исходные данные, давайте посмотрим. Они представлены в виде вот такого файла. Вам видно должно быть, да? Всё нормально сейчас. Значит, что за файл? Значит, здесь строчки — это наблюдения, колонки — это э-э переменные, можно так сказать, если использовать терминологию SPSS, статистика вот эта IBM. Э-э эти переменные делятся на две категории: зависимые и независимые. То есть переменные, которые являются факторами, действующими на объект моделирования, и переменные, которые являются результатами влияния этих факторов.

Запуск системы Эйдос (локально)

Сейчас я открою... Да, да, запущу саму систему сначала. Вот Start Eidos — это как раз модуль, который обновление обеспечивает. Вот она пишет, что диагностика показывает, что нет FTP соединения с сервером. Естественно.

9. Режим 6.4: Формализация и обработка данных

Обзор режима

Вот. И, значит, мы видим здесь. Значит, здесь есть режим 6.4, который описывает, что, какие функции выполняет система Эйдос. Она преобразует исходные данные в информацию, а информацию в знания.

Данные -> Информация (Причинно-следственные связи)

Каким образом? В соответствии с концепцией смысла Шенка и Абельсона, э-э данные преобразуются в информацию, если мы в этих данных обнаруживаем причинно-следственные зависимости. Вернее, обнаруживаем причинно-следственные зависимости в той предметной области, которая описывается этими данными. Если мы знаем эти причинно-следственные зависимости, то есть мы понимаем смысл данных. А да, согласно концепции смысла Абельсона, концепции смысла Шенка Абельсона, смысл — это знание причинно-следственных зависимостей. То есть тогда мы понимаем, что у нас является причинами, что последствиями, и как причины влияют на последствия.

Информация -> Знания (Поддержка принятия решений)

И потом, если мы знания этих причинно-следственных зависимостей используем для решения задач поддержки принятия решений, то есть для управления, тогда эти эта информация превращается в знания.

Адаптивность системы Эйдос

Система Эйдос всё это обеспечивает. То есть она обеспечивает и преобразование исходных данных в информацию путём поиска в них причинно-следственных зависимостей, и использование этой информации для принятия решений, для управления. То есть эта система может быть использована в составе систем управления для принятия решений. Причём система управления адаптивная, которые могут использовать в соответствии с принципом дуальности управления Александра Фельдбаума, могут использовать результаты управления для улучшения модели, которая использовалась для принятия решений. То есть интеллектуальные системы работают не на основе одной какой-то модели предобученной, созданной при конструировании системы, как раньше, но и в самом процессе управления могут улучшать эту модель с учётом результатов управления, и могут её также поддерживать адекватной в случае изменения этого объекта управления и потери адекватности предыдущей модели. То есть когда модель, объект моделирования проходит точку бифуркации, изменяются основные закономерности, которым он подчиняется, объект моделирования, или мы выходим за пределы генеральной совокупности, где действуют другие закономерности, то адаптивные системы это выдерживают. Неадаптивная система теряет адекватность и перестаёт э-э, ну, в общем, теряет эффективность.

Этапы в режиме 6.4: Формализация, кодирование, обучение

И вот здесь вот в этом режиме 6.4 мы видим, что сначала мы осуществляем формализацию предметной области. На этом этапе мы разрабатываем э-э как раз э-э классификационные описательные шкалы и градации и используем исходные данные, э-э то есть используем эти класс, эти шкалы и градации для кодирования исходных данных. В результате чего у нас получается обучающая выборка.

Анализ примера данных (предприятия)

Вот мы смотрим на эти исходные данные. У нас они представляют собой наблюдения разных предприятий, у которых, вам видно вообще, как вот здесь? Да, всё хорошо видно. Спасибо. Вполне, да? Вот. Сейчас посмотрим, может, как-то... Вот. То есть у нас вот эти колоночки, не без фона которые, это колоночки, соответствующие интегрированным показателям бухгалтерского учёта в соответствии с монографией Анны Владимировны, которая по результатам защиты кандидатской диссертации была опубликована. Эти исходные данные взяты полностью один к одному из этой диссертации. Прямо вот, э-э, вот они тут эти данные, по-моему. Да. Вот эти перечисляются эти показатели, какой их смысл. И дальше идёт обработка.

Классификация предприятий (Кризис/Норма)

Вот, значит, здесь в качестве исходных данных использованы данные по ряду предприятий, которые по экспертным оценкам были в кризисном состоянии, близком к состоянию банкротства, и также в нормальном состоянии, далёком от состояния банкротства. А также, да, из этических соображений эти предприятия здесь не названы. Просто написано, что тридцать первое, тридцатое, двадцать девятое предприятие там, первое, второе, тринадцатое и так далее.

Классификационные и описательные шкалы

А вот эта шкала, которая выделена фоном, это классификационная шкала. Их может быть несколько. И, как вы видите, описательных шкал здесь довольно много, ну, сравнительно. И одна классификационная. Значит, вообще я, когда рассказываю об этом, я всегда говорю так, что классификационные шкалы отражают результаты влияния факторов в двух вариантах это влияние можно формализовать, описывать: в натуральном выражении и в стоимостном выражении. В натуральном выражении обычно это количество, качество продукции, в стоимостном выражении обычно это прибыль и рентабельность. Эти шкалы могут быть текстового типа и числового типа. Если в шкале есть хотя бы одно значение текстового типа, то она все остальные значения тоже рассматриваются как текстовые. В системе есть возможность рассматривать в текстовых шкалах пробелы как отсутствие данных, это просто опция в интерфейсе автоматизированном ввода данных. А в числовых шкалах мы можем ноль рассматривать как отсутствие данных, а можем рассматривать как значащее значение, как имеющее смысл значение. И также и пробелы.

Пример данных: Предприятия и их классы

И вот здесь вот у нас есть и ряд предприятий указано конкретно. Значит, вот эти вот категории, классы "кризис" и "норма" — это обобщённые категории, обобщённые классы. А вот дальше идут классы, созданные на основе одного единственного примера, одного предприятия, относящегося к классу, который соответствует его названию. Можно считать, что вот это вот распознаваемая выборка или тестовая выборка. А вот эта вот до сорока наблюдений — это у нас тренировочная выборка или обучающая выборка.

Варианты использования данных (обучение/распознавание)

Значит, но есть разные варианты использования интерфейса. Можно его использовать для создания, для формализации предметной области, создания классификационных описательных шкал и градаций и обучающей выборки и создания впервые модели или адаптации модели объекта моделирования. А можно использовать уже существующую модель для формирования распознаваемой выборки. И есть соответствующие там опции, режимы в этих интерфейсах. Сейчас я покажу. А можно прямо взять и эту распознаваемую выборку включить в обучающую, но просто не указывать классы вот эти. То есть можно здесь вот всё стереть и останется, первая колоночка не является шкалой, она содержит информацию об источнике данных. И тогда будет проведено распознавание вот этих вот наблюдений и классификация по тем классам, которые у нас были указаны. А если оставить названия, указать названия этих вот наблюдений как классификационную шкалу и включить, тогда можно, используя кластерный анализ и результаты идентификации, сравнивать эти предприятия, которые здесь вот у нас перечислены, с обобщёнными образами и друг с другом. То есть есть три варианта таким образом. Вообще их не указывать, а только в распознаваемой выборке включить. Включить их прямо целиком в обучающую выборку и использовать наравне с обобщёнными классами. И вообще их не указывать, а включить в обучающую выборку, и будет просто система их относить, не будет формировать на основе них классов, а просто будет их использовать для идентификации на этапе обучения для при проверке достоверности модели.

Загрузка данных в систему

Вот. Значит, теперь мы в самой системе входим в программный интерфейс. Значит, да, мы вот здесь вот можем на лабораторных работах это изучать всё. А можем взять и... Сейчас я этому удалю то, что там приложения какие-то есть. Система является оболочкой, позволяющей создавать интеллектуальные приложения и runtime-системой, в которой, в среде которой, среда исполнения этих интеллектуальных приложений. Эти приложения могут использоваться, причём в адаптивном режиме, потому что инструментарий в наших руках их создания и, соответственно, их модификации. А можем создавать сами, не используя учебные приложения, а сами можем создавать интеллектуальные приложения на основе своих данных.

Интерфейс 2.3.2: Ввод исходных данных

Вот сейчас я вышел на подсистему 2.3.2, в которой мы видим все программные интерфейсы вот здесь. И различные служебные режимы, работа с этими интерфейсами, и специализированные программные интерфейсы, которые вот я разработал для участия в чемпионате России в открытом по искусственному интеллекту. И в этом чемпионате я по оценке экспертного сообщества я в двух номинациях занял первое место, в одной второе. Вот здесь вот эти как раз интерфейсы, я их оставил на память. Есть соответствующие публикации, где эти сертификаты, всё там есть. Вот сейчас мы выходим на интерфейс 2.3.2, автоматизированный программный интерфейс ввода исходных данных.

Параметры интерфейса (из файла .arch)

Причём я скопировал из папочки приложения, скопировал файл 2322.arch, который и представляет собой файл вот этих параметров, которые вводятся в этот программный интерфейс. Что это за параметры? Если я держу клавишу Shift и двигаю вправо э-э переключением на клавиатуре стрелочкой, э-э то мы видим вот здесь вот в строке формирования, то есть формируется формула соответствующая. То есть, по сути дела, мы считаем эти колоночки. И вот мы посчитали, что у нас всего 17 колоночек. Вторая колоночка у нас является классификационной, с третьей по семнадцатую являются описательными шкалами, то есть факторами, которые описывают факторы.

Настройка шкал (классификационные/описательные)

Вот, значит, мы видим здесь, что здесь заданы классификационная шкала со второй по вторую, а описательная с третьей по семнадцатую. Можно было сделать наоборот, можно было классификационные шкалы разместить после описательных, кому как удобно. Значит, или как данные были представлены исходные. Но для системы это совершенно всё равно. Но, значит, здесь указан тип исходных... Главное, они должны идти подряд, чтобы можно было вот так указать. Можно также указывать, допустим, вторая классификационная шкала, а описательная, скажем, вот с какой-нибудь там с четырнадцатой по семнадцатую, к примеру. И или классификационные вот эти задать, а это описательные. То есть тут самые разные варианты можно использовать.

Завершение этапа формализации

Значит, теперь этот Help я убираю. Значит, здесь дальше мы создаём эти шкалы и градации, обучающую выборку, потом модели, определяем достоверную модель, если она есть, то решаем в ней задачи, которые классифицирую условно так несколько на задачи идентификации и прогнозирования, задачи принятия решений и задачи исследования моделирования предметной области. Но они, почему я сказал достаточно условно это разделение? Потому что они пересекаются, в части используются задачи прогнозирования принятия решений, они как прямая и обратная задача являются, рассматриваются. При задаче принятия решений в развитом варианте используются результаты решения задачи исследования. Поэтому несколько условно это разделение.

Дополнительные опции обработки данных

Значит, дальше вот здесь вот мы как раз то, о чём я говорил. Нули и пробелы считаем отсутствием данных. И здесь ещё много разных опций, которые я сейчас не буду, может быть, рассказывать, это детали различные. Ну я так скажу, что они есть. Значит, и числовые шкалы мы можем разделять на диапазоны числовые, интервальные значения, потом будем обрабатывать двумя способами: равные интервалы с разным числом наблюдений или разные интервалы с примерно одинаковым числом наблюдений. Примерно, потому что в целых числах. То есть полтора землекопа не бывает, там или два, или три. Поэтому вот эти с равным числом наблюдений, но в целых числах, то есть плюс один или минус один. И мы можем либо формализовать предметную область, то есть создать шкалы и градации обучающую выборку, либо использовать уже существующую модель для создания обучающей выборки и решения задач. Мы можем применять сценарный метод системно-когнитивного анализа, то есть как признаки рассматривать предыдущую динамику изменения параметров вот этих, которые у нас без жёлтого фона факторов, и на заданную глубину предыстории. И также как результат влияния факторов мы можем рассматривать будущую динамику изменения тех показателей, которые у нас описываются классификационными шкалами, тоже на заданный горизонт прогнозирования. То есть она может анализировать сценарий, не только сами точечные значения наблюдений, но и сценарии их изменения. И мы можем использовать э-э при равных интервалах, можем использовать текстовую интерпретацию значений. Это мы рассмотрим при анализе текстов. Здесь у нас есть возможность использовать как признаки э-э как значения факторов э-э слова, разделённые стандартными разделителями, как и в факторах, в описательных шкалах, так и в классах. Причём по-разному, разные варианты есть здесь. Но мы сейчас не будем это делать, это будем на при анализе текстов делать. И сейчас мы оставим все параметры так, как они были э-э в самой этом, в самом этом приложении выбраны.

Запуск обработки

Нажимаем О'кей. Происходит конвертация внешнего файла экселевского в внутренний формат данных системы, то есть тот формат баз данных стандарт, который используется в языке программирования.

Результаты предварительного анализа

И мы, значит, предварительный анализ проведён, мы видим, сколько здесь классификационных шкал, сколько описательных, сколько градаций у них и так далее. Вот в классификационных шкалах не обнаружено числовой шкалы, поэтому здесь не спрашивается, сколько числовых диапазонов. А в описательных шкалах обнаружены числовые шкалы, их там много, и поэтому возникает вопрос об числе диапазонов. И здесь задано два.

Сформированные классы и признаки

Значит, таким образом у нас сейчас получается 22 класса — это градации классификационных шкал, которые включают в себя вот это вот кризис, норма и названия предприятий конкретных. И 30 признаков — то есть это это градации описательных шкал. Ну, это низкое и высокое значение, грубо говоря, когда две градации.

Просмотр шкал и градаций

Мы можем посмотреть на то, что у нас получается в результате. Вот у нас вот такие классификационные шкалы получились. Они по алфавиту здесь рассортированы классы. Одна классификационная шкала. И там мы видим названия предприятий, а также видим вот кризис и норма. И описательные шкалы мы все видим здесь. И у нас диапазоны, на которые они разделены, диапазон значений каждой шкалы разный. И внутри этого диапазона разбита шкала на интервалы, которые разные по величине, но такие, что 29 наблюдений на диапазон. Что более чем достаточно для того, чтобы можно было говорить о том, что это статистически должно быть значимо, достаточно нам. Ну как считается так, что как минимум должно быть пять наблюдений. Ну здесь почти 30 наблюдений. Значит, это, в общем-то, говорит о том, что у нас есть основания утверждать, что те результаты, которые будут получены, они, ну, основаны на повторяемости явлений. Повторяемость — основная, основной критерий, который позволяет говорить о том, что мы имеем дело с какими-то закономерностями, а не со случайностями.

Просмотр обучающей выборки (закодированной)

И вот, значит, можно задавать другое количество, конечно, диапазонов, тогда нужно пересчитать шкалы и градации. И, значит, э-э, в принципе, подбора я вам сказал, основной принцип, что должно быть достаточно наблюдений для того, чтобы можно было говорить о закономерностях. При небольшом количестве данных, как у нас здесь сравнительно, значит, э-э, совету рекомендуется не очень много этих диапазонов брать. Значит, ну я могу вам привести пример простой. Если взять 100 диапазонов сейчас, то у нас многие диапазоны не будут иметь ни одного наблюдения. Вот. Поэтому мы должны понимать, что их должно быть меньше, чем наблюдений. Ну, причём несколько раз меньше. И есть ещё очень хорошая такая теорема Котельникова или Найквиста-Шеннона, ещё её за рубежом называют. Хотя Котельников на 15 лет раньше её вывел. Вот, и правда, не смог опубликовать работы, ему не давали публиковать её, когда ещё он был студенческий такой, в таком аспирантском возрасте, не давали ему публиковать. Ну это известная история. Вот. Теорема об отсчётах, связанная со спектрами сигналов, с частотой отсчётов. Вот. Значит, смысл такой, что э-э когда мы берём адаптивные интервалы, то они таким образом разбиваются вот системой Эйдос, в соответствии с теоремой Котельникова, чтобы у нас не было лишнего количества интервалов, но те, которые есть, давали максимально позволяли представить спектр сигнала.

Переход к синтезу моделей

Вот выходим мы на уже создание классификационных описательных шкал и градаций обучающей выборки. Мы видим здесь, что выполнены следующие работы: созданы, сформированы классификационные описательные шкалы и градации, и они использованы для кодирования исходных данных, в результате чего у нас получилась обучающая выборка. И теперь потом мы выполняем небольшие операции переиндексации для поиска. И дальше у нас получается э-э такие вот классификационные шкалы и градации.

10. Синтез и верификация моделей

Типы моделей в Эйдос

Вот. Если какие-то вопросы возникают, вы сразу говорите. Может быть, я тогда сразу буду давать пояснения. Значит, вот у нас кризисная класс кризис, класс норма и классы, соответствующие конкретным предприятиям. То есть, по сути дела, те, которые мы хотим оценить. И э-э описательные шкалы. Здесь в каждой колоночке вот этой, которая была без фона вот здесь, вот каждой колоночке соответствует у нас э-э шкала описательная, а справа мы видим градации этой шкалы. Не все у нас числовые. Как вот можно пролистать их, посмотреть. Если, значит, какие-то есть не числовые, то здесь сразу будет видно, что числа будут рассматриваться как тексты. Здесь будет много-много градаций в этой шкале. Это может означать, что там либо где-то вместо запятой точка указана, либо где-то там буковка какая-то есть, либо пробел там, где отсутствие данных, или чёрточки какие-нибудь, или Н/Е там, Н/Я, отсутствуют данных там и так далее. Вот мы смотрим обучающую выборку. Значит, здесь у нас вверху название объекта наблюдения. Внизу слева — коды классов в соответствии с классификационными шкалами. Внизу справа — коды признаков. Ну, то есть признаком мы как кратко называем, краткая такая терминология удобная. В качестве признаков мы рассматриваем градации описательных шкал. То есть либо это значение свойств, ну, допустим, шкала цвета, значение красной, либо, значит, это значение факторов, ну, например, там какой-то полив там интенсивный, например, там сильный там. И вот так вот мы видим, закодированы уже все исходные данные. Вот этот процесс кодирования осуществляется очень быстро. Э-э, то есть буквально там миллионы наблюдений, 800.000 наблюдений, где-то может за 7 минут закодировано будет вот так примерно. То есть кодирование идёт с очень большой скоростью.

Режим 3.5: Синтез и верификация

Когда этот интерфейс используется 2.3.2, то его можем также вот такой форме просмотреть уже обучающую выборку. Здесь мы видим, что, значит, вместо наименований и числовых, то есть текстовых наименований, и здесь норма, кризис и название предприятия, уже коды соответствующие указаны. А здесь вместо чисел указаны коды числовых диапазонов. Там, где пробел, это, значит, мы здесь имеем дело с отсутствием данных. То есть данные у нас без полных повторностей, частично фрагментированы, но степень фрагментации невысока. Дальше у нас по вот этой методике решения задач, у нас дальше идёт синтез моделей. Сначала создаются модели статистические, классическая матрица корреляционная или матрица абсолютных частот, которая абсолютно во всех системах статистических и, наверное, интеллектуальных, в большинстве систем создаётся. Потом на основе неё рассчитываются две модели статистические, которые я называю матрица условных и безусловных процентных распределений. То есть от абсолютных частот, частот мы переходим к относительным частотам. А потом переходим к моделям знаний, где используются уже э-э частные критерии, которые представляют собой результаты сравнения абсолютных и относительных частот. Абсолютные частоты мы фактически сравниваем с теоретическими, это хи-квадрат классический критерий Карла Пирсона, замечательный. А вот эти вот условные и безусловные относительные частоты мы сравниваем друг с другом, то есть условные сравниваем с безусловными. Я модели поясню чуть попозже, как они получаются.

Конфигуратор моделей и критерии

Значит, в системе реализован конфигуратор системно-когнитивных моделей. То есть, что такое конфигуратор? Это означает, что больше нет способов сравнения э-э ну, неизвестно в науке больше способов сравнения абсолютных частот э-э фактических с теоретическими, относительных частот условных с безусловными, чем то, э-э чем те, которые реализованы в системе Эйдос. То есть все, которые известны, все реализованы. Ну, немножко я кривлю душой, на самом деле мне ещё два известно, но я их не реализовал. Они мне не нравятся, если честно скажу. Значит, ну вот они это, ну, скажу, когда будем об этом говорить.

Запуск синтеза и верификации

Значит, вот теперь мы создаём модели таким образом. Для этого при... там много есть разных режимов, но наиболее таким употребительным является режим 3.5. Здесь мы видим что? Что создаются вот эти модели статистические, модели системно-когнитивные, модели знаний. При этом может только синтез модели осуществляться, синтез и только верификация моделей, синтез и верификация.

Верификация и бутстреп-подход

Верификация — это проверка достоверности модели путём решения задачи идентификации на объектах обучающей выборки. Если обучающая выборка очень большая, то процесс идентификации может быть очень э-э затянутым во времени, потому что для того, чтобы идентифицировать объект, то есть наблюдение, этот объект сравнивается со всеми классами по всем своим признакам. То есть это огромный объём вычислений. Поэтому при очень большом объёме наблюдений, там когда их там тысячи, десятки, сотни тысяч, система позволяет где-то примерно по оценкам около 12 млн наблюдений обрабатывать. То есть в Экселе столько мы представить не можем, но она может и с CSV файлами работать и может обрабатывать подобные объёмы наблюдений. То есть приближающиеся к большим данным, я бы сказал так. Ну, строго говоря, это ещё не большие данные, но довольно большие, скажем так. Вот. Э-э есть возможность использовать графическую процессоры для расчётов. Это в том случае, если у нас видеокарта есть, которая поддерживает OpenGL язык параллельного программирования, на котором программируются вот эти графические процессоры. Для этого должно, чтобы этот язык поддерживался, видеокарта должна быть либо то есть на чипсете Nvidia, либо просто Nvidia, либо GeForce, значит, ну в общем, на чипсете Nvidia должна быть. Используется, то есть система поддерживает также бутстреп-подход. То есть если мы не можем всю все всю обучающую выборку использовать для проверки достоверности модели из-за её очень большой величины, то можно очень быстро создать модель, это процесс довольно быстрый, даже на центральном процессоре. Вот, и потом произвести верификацию не на всём объёме выборки, а на каком-то подмножестве этой выборки. То есть мы выборку исходную разбиваем на две части. Одну часть используем для создания модели, а другую для проверки достоверности. Мы такую задачу решали, когда мы 30.000 респондентов имели базу данных с коллегами. Э-э, значит, убрали все наблюдения, по которым меньше было, то есть все классы убрали, по которым меньше тысячи наблюдений было. Это в геофизике мы такие тоже задачи решали. То есть тысячи наблюдений на класс, не менее. Получается классов тогда поменьше, уже не 11.000 классов получалось там, а уже 380. Вот мы для этих 300, 380 очень хорошо представленных классов решали задачу следующим образом: мы на 25.000 наблюдений создали модель, а на 5.000 их жёстко в режиме таком с пристрастием проверяли на достоверность, используя именно бутстреп-подход. Здесь можно случайным образом выбирать заданное число объектов или там какой-то диапазон или каким-то образом их выбирать. Их можно либо удалять, либо не удалять из обучающей выборки. То есть считается, что чище будет проверка, если удалять. Но я вам скажу, что моё мнение такое, что когда очень много наблюдений, то никакой никакого влияния эти наблюдения уже не оказывают, каждое отдельное наблюдение мало влияния оказывает на модель. То есть если там десятки тысяч наблюдений, от того, что мы удалим какое-то одно наблюдение, ничего практически не меняется в модели. То есть можно не удалять, по фактически это значит.

Внутренний критерий достоверности

Ничего от этого не меняется. Так вот, э-э если мы используем бутстреп-подход, то мы берём, как делаем? Берём не очень большое число объектов, распознаём, только верификацию. Система замеряет время в миллисекундах, которое на это ушло. Строим пропорцию простую, что для того, чтобы за ночь система посчитала, достаточно, значит, вернее, нужно нам задать не 100, а, скажем, 80 там тысяч наблюдений. Задаём, запускаем, идём спать, утром приходим, готова проверка достоверности на очень большом подмножестве э-э исходных данных. Здесь вот есть опция интересная. В системе есть внутренний критерий достоверности результатов идентификации, внутренний, адекватный критерий. Он неоднократно проверялся и, в общем, убедительный очень критерий. Сейчас я вам про него расскажу в двух словах потом. И система сама знает, насколько достоверно её решение идентификации. И может убрать недостоверные решения. То есть я могу здесь вот задать 10, например, вместо 100, тогда она оставит 10% наиболее достоверных решений. Это может иметь смысл при очень больших данных. А так обычно, ну и влияет на скорость обработки информации, на размерность баз данных и так далее.

Запуск синтеза и верификации (пример)

Значит, мы сейчас э-э системой э-э создаём модель, которая составляет 3 тысячных процента, послушайте, уважаемые коллеги, 3 тысячных процента от того, что может обрабатывать система Эйдос. То есть она может обрабатывать значительно модели значительно большей достоверности.

11. Результаты идентификации и их интерпретация

Процесс распознавания

Значит, сейчас мы на графическом процессоре проводим распознавание 59 объектов у нас наблюдений. И э-э вот с такой скоростью примерно процесс идёт. Значит, здесь идёт э-э отображение стадии исполнения и э-э одновременно э-э прогнозирование времени исполнения. И постоянно этот прогноз уточняется, потому что здесь много разных операций осуществляется: сначала синтез модели, потом их использование для проверки, потом они проверяются на достоверность путём решения задачи идентификации. Для этого можно и другие задачи использовать, но проще всего задачу идентификации для этого использовать. И поэтому прогноз постоянно корректируется и уточняется.

Выходные формы

Вот мы видим, что за 2 секунды проведено, проведена идентификация. Получены, вот дальше вот эти вот сообщения — это сообщение о том, что были выданы, э-э что были сформированы э-э выходные формы. Их на данный момент 11 выходных форм по результатам решения задачи идентификации, разные выходные формы.

Анализ результатов (пример)

И вот некоторые из этих форм можно посмотреть. Одну я вам уже показал до этого, две даже показал. Вот мы видим, что вот это первое предприятие кризисное, очень похоже на обобщённый образ э-э кризисных предприятий, а также похоже на АвтоВАЗ в первом квартале 2004 года, а также на какую-то там э-э ОШЭ, там, я не знаю, что это такое, УСИ, Аэрофлот в первом квартале двадцать четвёртого года, Дорсервис, АвтоВАЗ и так далее. И вот, значит, мы каждое вот это наблюдение можем посмотреть. То есть мы можем предположить, что это и есть эти предприятия, что они и использованы для создания этих образов. Вот. А вот нормальные, которые как нормальные наблюдения. И здесь у нас наблюдаются странные вещи. Вот, допустим, когда с кризисом, то вот мы здесь видим вверху кризис, видите, вот. И поэтому интегрально... Да, это окно правое, окна, два окна, верхнее и нижнее — это с разными интегральными критериями. Сейчас я про них расскажу в двух словах, когда буду про модели рассказывать. И вот здесь вот мы когда по кризису э-э пробегаем по кризисным наблюдениям, то мы видим везде у нас плюс. То есть, ну вот. А вот когда мы начинаем норму просматривать, мы видим раз, что у нас оказывается вот это предприятие 20, которое экспертами отнесено к нормальным, объективно по своим показателям, э-э бухучёта, которые используются в соответствии с подходом для управления. Вот, для принятия решений. Оно оказывается похоже на кризисное больше, чем на нормальное. На нормальное оно не похоже, -6,5%. То есть это, возможно, является неточностью в исходных данных. Мы такое, по-моему, только одно заметили тут.

Уровни сходства и нечёткость

И дальше идут предприятия конкретные. И вот мы видим здесь, что система их оценивает эти предприятия, насколько они похожи на что. Вот, допустим, Аэрофлот-1, э-э 2000, первый квартал 2004 года, он больше похож на кризис, вот на 8%, 9, чем на норму. Норма с минусом здесь стоит. И так вот мы по всем предприятиям видим результаты их идентификации. По сути дела, мы видим сходство их с другими предприятиями и сходство с обобщёнными образами, созданными на основе тех предприятий, которые экспертами отнесены э-э к нормальным и кризисным по состоянию своему. И вот здесь мы видим кризис, сами эти кризисные наблюдения. А также вот, смотрите, прямо вот идут у нас конкретные предприятия и также наблюдения. И то же самое касается нормы. То есть мы, по сути дела, провели уже оценку предприятий, решили задачу идентификации. Из этих вот... Ну я уже говорил, что это можно сделать разными способами. Мы это сделали в наиболее достоверной модели. И вот эту задачу решили, задачу идентификации.

Связь с прогнозированием и принятием решений

Задача прогнозирования мало чем отличается от задачи идентификации. Математически она вообще не отличается, на мой взгляд. А отличается она тем, каким способом мы интерпретируем результаты. Если мы при идентификации говорим о свойствах, их значениях и обобщённым категориям, которым сейчас относятся объекты с такими свойствами и их значениями, с признаками, то при распознавании, это при распознавании идентификации и диагностики, то при прогнозировании мы говорим так: какие значения факторов каким образом влияют на объект моделирования, на его будущее поведение и переход в будущее те или иные будущие состояния, соответствующие классам. Ну то есть разница в том, что при идентификации, э-э диагностике и распознавании мы используем трёхмерное пространство время. А при э-э прогнозировании мы используем четырёхмерное пространство время фактически. В этом смысле можно сказать так, что прогнозирование — это та же самая идентификация, только не в трёхмерном пространстве, а в четырёхмерном пространстве времени.

Задача принятия решений (обратная задача)

Теперь можно посмотреть, что у нас там связано с принятием решений. Принятие решений тесно связано с прогнозированием. Даже можно пытаться решать задачу управления путём прогнозирования, но это не получится при больших, при большом числе факторов из-за комбинаторного взрыва, большого числа сочетаний значений факторов. Но при небольшом можно даже это пытаться делать. Но вообще-то говоря, вообще говоря, при прогнозировании мы по значению факторов определяем, какое будет будущее состояние объекта моделирования. А при принятии решения мы наоборот, по будущему состоянию определяем, какие значения факторов необходимы для перевода объекта в это будущее состояние. То есть получается что? При прогнозировании мы по значению факторов определяем будущее состояние, а при принятии решения наоборот, по будущему состоянию, целевому, желательно, определяем, какие значения факторов обуславливают это состояние. То есть это прямая и обратная задача.

Свот-анализ и рекомендации

Поэтому, значит, мы решаем эту... Да, кстати, насчёт того, что прогнозирование можно использовать для управления. Значит, мы вот здесь вот в режиме 2.2, где смотрим на шкалы, мы можем нажать помощь, и мы видим, что у нас при данном количестве описательных шкал и их значений, э-э получается 32.768 сочетаний факторов, что примерно, э-э если на каждое сочетание тратить 1 секунду, то примерно э-э полдня уйдёт на прогнозирование различных результатов э-э действия различных сочетаний значений факторов. А вот если мы возьмём режим 4.4.8, э-э это решение обратной задачи прогнозирования, то мы увидим просто, какие значения факторов обуславливают то или иное состояние. Вот состояние норма вверху выбираем. Вот здесь уже не надо нам текущие модели, здесь задача решается быстро, поэтому перечислены все модели. Берём, клацаем по наиболее э-э достоверной модели и видим с весовыми коэффициентами, как влияет то или иное значение фактора на переход объекта в это состояние, норма. И мы можем построить, это я называю свот-анализом, действуя, так сказать, э-э традиционным ключом, как в экономике, факторы делятся на внутренние и внешние, способствующие и препятствующие достижению некоторого состояния. Вот если мы выбираем здесь состояние, то слева мы видим способствующие факторы в порядке убывания степени их влияния на переход объекта моделирования в это состояние, а справа препятствующие. И можем это выразить в графической форме. Да, и здесь есть и таблицы, которые открываются в Экселе. Кстати, стандарт баз данных системы Эйдос, он такой, что Excel открывает этот эти базы данных. Это dBase формат, это Excel 2 фактически, который был когда-то, когда было ещё Windows не операционная система, а под ДОСом программы были. И вот мы берём и строим свод-диаграмму. Здесь мы видим факторы наиболее сильно влияющие на переход объекта в это состояние, здесь наиболее сильно препятствующие. И вот мы видим максимальный коэффициент покрытия запасов, максимальный коэффициент быстрой ликвидности, вот весовые коэффициенты здесь указаны, высокая автономность собственных средств, высокая обеспеченность собственными запасами, оборотными средствами, общая высокая рентабельность, низкий коэффициент финансовой зависимости и низкий индекс постоянного актива. Ну вот те, кто смысл этих коэффициентов знает, они сразу скажут: "Ну да, конечно". Вот. И, в общем, мы это видим, что система это узнала. То есть она получила эти знания из данных о том, какие были показатели у предприятий.

Фильтрация и детализация анализа

Здесь есть аппарат определённый. Мы можем, сейчас это особого интереса не представляет, но вообще-то вот здесь есть возможность отфильтровать. Вот, допустим, здесь мы сейчас берём и фильтр устанавливаем по определённому фактору и видим, как влияют градации этого фактора на тот или иной результат. Вот. И фактор у нас был F2. Вот находим мы фактор F2, включаем фильтр, получаем, что высокий коэффициент автономности собственных средств — это хорошо, это характерно для нормальных предприятий, а низкий — это характерно для предкризисного состояния. Когда очень много градаций у факторов, тогда этот режим интересен. Когда мы видим, э-э устанавливаем фильтры. И мы можем так вот любой фактор посмотреть, вернее, любой класс посмотреть, как он сформировался, в чём сущность этого класса. Вот. И можем здесь и предприятия смотреть, конечно. Вот. Для что наиболее характерно для Лукойла, например, в первом квартале 2004 года? Наиболее для него характерно что? Низкая оборачиваемость активов, высокий коэффициент покрытия запасов, ну и так далее. Это сравнение с другими предприятиями и с обобщёнными образами.

Практическое применение: Рекомендации и ограничения

То есть мы, по сути дела, провели уже оценку предприятий, решили задачу идентификации. Из этих вот... Но я сказал, что это можно сделать разными способами. Мы это сделали в наиболее достоверной модели. И вот эту задачу решили, задачу идентификации. Задачи принятия решений. То есть если мы хотим выработать рекомендации для предприятия, то мы говорим так: так, вот у нас наши данные говорят о том, обобщённые наши данные, результат анализа говорят о том, что желательными являются вот такие значения вот этих вот э-э коэффициентов. Вот мы рекомендуем вам подумать о том, чтобы повысить э-э коэффициент покрытия запасов, коэффициент быстрой ликвидности, автономность собственных средств. Надо вам повышать это, ребята, эти показатели, значения их повышать. Тогда вы нормализуете состояние своего предприятия. Но э-э заказчик может сказать так: "А я не могу этого сделать. Вот этот коэффициент я не могу взять и повысить. У меня нет для этого там средств каких-то технологических". Вот. Поэтому, значит, что мне делать? Тогда, значит, что мне делать, на этот вопрос отвечает уже следующие результаты анализа, но сейчас я потом вам уже результаты исследования предметной области.

12. Исследование предметной области и когнитивные функции

Кластерный анализ и дендрограммы

Сейчас сначала вам покажу, что мы можем ещё посмотреть, как влияет любой фактор. Берём любой фактор какой-нибудь и смотрим, как он влияет на предприятие. Вот он, вот это значение, высокое коэффициент покрытия запасов, оно э-э положительно влияет на переход в состояние нормы. Вот, э-э препятствует переходу в кризисное состояние, характерно для этих предприятий, а для тех не характерно. То есть мы знаем, как влияет каждый фактор. Это называется семантический э-э портрет э-э значения фактора. То есть смысловое его, смысловая характеристика этого значения фактора. Также мы можем посмотреть в разных формах это. Ну, допустим, посмотреть на нелокальные нейроны. Что у нас характерно для нормальных предприятий? Посмотрим, что у нас характерно. В стиле нейросетей. Мы видим сам, тело нейрона — это у нас соответствует классу, а рецепторы соответствуют э-э значениям показателей финансовых. Э-э и красные — это активирующие этот нейрон, синие — тормозящие. Э-э надписи на рецепторах обозначают смысл этого э-э шкала фактора и его градации. А толщина линии соответствует степени влияния. Если очень много у нас э-э рецепторов, бывает их там тысячи, тогда здесь ничего не поймёшь. Тогда вот здесь можно задать ограничение, сколько показывать рецепторов. Ну, допустим, 24 показывать. Тогда будет видно, соответственно, поменьше их. Вот, или 12. Или показать только те, которые наиболее сильно влияют по модулю информативности, по самому информативности, сортировки различные тут есть и так далее.

Дендрограммы и кластеризация

И э-э также есть, это то же самое, что информационный портрет класса. И тут ещё есть возможность все нейроны одновременно визуализировать. При небольшом количестве классов это осмысленно и рецепторов. Когда их очень много, то здесь есть различные средства, позволяющие ограничить и посмотреть не на всю модель, а на часть модели. Мы видим здесь, что показано примерно половина созданной модели. Мы видим, что наиболее, да, вот эти красные линии — это детерминация, то есть активирующие связи, синие — это э-э тормозящие связи. Мы видим здесь у нас два класса норма и кризис, которые сильно обусловлены значениями факторов. И видим классы, соответствующие предприятиям, которые слабо обусловлены по сравнению с этими классами. Вот. То есть это можно, это называется круговая когнитивная диаграмма. Есть э-э модели представления знаний, которые прямо и представляются вот такой круговой когнитивной диаграммой. Это сетевые модели представления знаний. То есть прямо строится вот такая диаграмма и рисуются вот эти линии на основе экспертных оценок. И в результате считается, что создана модель знания. Ну, действительно, она содержит знания о том, насколько сходны и отличаются вот эти состояния, соответствующие классам объекта моделирования, по их системе детерминации, то есть по тем значениям факторов, которые их обуславливают. Но красивая форма, вот, но, честно сказать, мне кажется, что более информативной является дендрограмма. Вот. В дендрограмме мы можем задать параметры до 4К. Особенно по оси X это роли особой не играет, а по оси Y вот здесь вот играет роль. Э-э из практики я могу сказать, из опыта, что где-то до 400 объектов ещё можно рассмотреть, что там такое, если бы здесь вот задать 4.000 пикселей по вертикали. Но здесь тогда желательно крупный шрифт задать. Вот. Ну сейчас я не буду этого делать, задам только крупный шрифт. И э-э модель F3 мы смотрим формирование дендрограммы агломеративной и её построение. Значит, в системе Эйдос особенность есть такая, что она работает с длинными наименованиями объектов наблюдения, классов, признаков. И они все помещаются, короче говоря, на формах нормально. Вот это вот название вверху, это можно в режиме 1.3 поменять, будет написано то, что нужно здесь. Мы видим, что у нас есть два э-э полюса конструкта: синие и красные э-э кластеры. Значит, э-э синие группируются вокруг кризиса, одиннадцатый класс. Красные группируются вокруг нормы, шестнадцатый класс. Мы видим степень сходства и различия. Ось X — это шкала различий между различными классами. То есть это решение задачи, по сути дела, э-э определения степени риска или или, скажем так, привлекательности для кредитования э-э тех или иных предприятий. Парадоксально, я хочу отметить этот момент, этот парадоксальный момент. В кредитовании больше всего нуждаются предприятия, которые находятся в кризисном состоянии, предкризисном. Они могут выйти из этого состояния именно путём инвестиций каких-то, кредитов. Э-э, замена, значит, оборудования, изменение технологии, реструктуризация предприятия и так далее. Всё это требует инвестиций. Но инвесторы не хотят давать инвестиции таким предприятиям, которые находятся в предкризисном состоянии, потому что высок риск невозврата инвестиций. То есть как раз тем, кому они нужны, им как раз и не дают. Потому что инвесторы заинтересованы в возврате инвестиций с прибылью, э-э с процентом определённым, а не в улучшении реальной ситуации предприятий. Поэтому подходы к принятию решений в инвестировании, они могут отличаться у инвестора одна точка зрения, а у получателя инвестиций совершенно противоположная обычно точка зрения. Вот если мы озабочены улучшением ситуации, то тогда мы должны принимать решения таким образом, чтобы вытащить предприятия предкризисные в нормальные. А если, значит, мы заинтересованы в собственной выручке, получении от инвестиционной деятельности, тогда мы должны давать портфель кредитный формировать таким образом, чтобы давать кредиты нормальным предприятиям, которые устойчивы и наверняка вернут наши кредиты с процентами.

Когнитивные функции

Вот. Поэтому, значит, есть теория в интеллектуальных технологиях, широко применяются интеллектуальные технологии в теории портфеля инвестиций с точки зрения инвестора. А вот с точки зрения управления и повышения эффективности в реальной предметной области, обычно мало кто рассматривает эти процессы. И у меня очень много работ в этой области, называется э-э рассмотрение инвестиций как фактора управления с целью перехода объекта моделирования, объекта управления, перевода объекта управления в заданное целевое состояние. Ну и как классическая форма, значит, э-э с увеличением межкластерного расстояния, сначала оно меняется несущественно, когда в начале кластеры формируются. И в самом конце уже ничего не удаётся сделать, как не объединить только полюса конструкта. Там резко скачет э-э различие. Вот эта форма, аннотация кластерная, когда вот так вот скобочками обозначены классы, которые объединены, это мною придумано, я нигде не видел ничего похожего. Значит, теперь смотрите, э-э такой момент интересный. Если мы нарисовали такую диаграмму, и у нас там что-то налепилось там одно на другое, непонятно, что там изображено, то тогда есть возможность перерисовать без перерасчёта. В параметрах мы меняем э-э параметры визуализации и рисуем сразу дендрограмму без перерасчёта. Дело в том, что при больших размерностях моделей может довольно э-э существенно, то есть довольно длительным быть процесс вот этот формирования дендрограмм. Так, сейчас я попытаюсь вам это скинуть. Вот. Ну, остальные я не скидывал, потому что они на экране были, то есть видно было. Вот. И здесь вот есть вот статья о когнитивном, э-э когнитивной кластеризации. Ссылочка на статью когнитивной кластеризации.

Отличие когнитивной кластеризации от классической

Почему я такой термин предложил когнитивная кластеризация? Потому что есть много разных... Я подверг критике методы кластеризации, меры расстояний, сказал о том, что, как правило, эти вот меры расстояния, которые используются, и в том числе в SPSS, ну там, правда, есть различные меры расстояния, и есть очень хорошие, такие как вот косинусное расстояние или межвекторное расстояние. Значит, э-э я считаю, что мера Эвклида — это, в общем-то, неудачная мера. Хотя она очень наглядная, естественная, понятная людям, но дело в том, что она работает тогда, когда оси когнитивного пространства э-э ортонормированы. Ну, то есть теорема Пифагора выполняется, грубо говоря, для этих осей. А, как правило, этого не происходит, не наблюдается. То есть обычно факторы взаимозависимы, э-э классы взаимозависимы. Я эти проблемы сформулировал вот в этой форме откровенной. Данные должны быть нормальными, должны быть сбалансированными. Тогда всё это хорошо работает. Если же это не так, тогда получается некорректные результаты. Так вот, модели системы Эйдос, они обеспечивают именно э-э высокое качество модели, адекватность в случае несбалансированных, неортонормированных, не э-э зашумлённых и фрагментированных исходных данных. То есть тогда именно, когда обычные статистические системы не работают. Ну, они-то работают как программа, но дают результаты, которые, как говорится, на голову не оденешь. И вот, значит, э-э я использую комбинацию системы Эйдос с другими системами. То есть я другие системы использую. Ну, допустим, кластерный анализ я провожу на больших размерностях, но не матриц абсолютных частот, а матриц информативности или модели хи-квадрат, которые формируются системой Эйдос и сделаны таким образом, чтобы их можно было легко э-э обработать в системе SPSS. Это предусмотрено. Есть режимы в системе Эйдос, это обеспечивающие удобство этой обработки. То есть там прямо название переменных, всё там включено, понимаете? Интерфейс через CSV файлы, CSV, Excel и туда в SPSS. И, пожалуйста, обработка идёт прекрасно. Получаются очень хорошие результаты.

Принцип когнитивной кластеризации

Вот в этом особенность системы Эйдос, что она делает нечто такое, что системы статистические вообще не делают. Зачем они не предназначены, ничего там вообще нету. Ну, по крайней мере, раньше так было. Сейчас-то они, ясное дело, интеллектуализируются, нейронные сети в SPSS включили. То есть сейчас, конечно, э-э меняется ситуация, но дело в том, что система Эйдос существовала и 20 лет назад, понимаете, патенты на неё были. И больше, 30 лет назад она существовала и решала эти задачи с таким функционалом она была, что видно. То есть она намного опережала, ну, свою по своей идеологии, намного опережала то время, когда она была создана. А сейчас это популярная система с открытым исходным кодом, используемая во всём мире, имеющая свои достоинства. Одним из других достоинств этой системы является наличие большого количества программных интерфейсов, которых вообще нет, скажем, в системе SPSS. То есть, то есть там их нет вообще, от слова вообще нет, понимаете? А это очень существенная часть, обеспечивающая, так сказать, удобство ввода исходных данных из сырого вида исходных данных, я имею в виду. В системе Эйдос это полностью автоматизировано для самых различных видов данных исходных.

Когнитивные функции (продолжение)

Ну, на этом тогда будем считать, что наше занятие закончено. Ещё один процент я вам рассказал того, что мог бы рассказать. Ну, остальные проценты, если будет принято решение продолжить эти семинары, то я расскажу тогда вживую остальное. Если же такого решения не будет принято, то вот есть э-э ссылки, которые я дал, большое количество монографий, статей, э-э уже имеющихся облачных Эйдос-приложений, которые можно изучить. Можно будет их изучить, когда сайту будет присвоен стандартный порт. Вот. И всё это, в общем-то, в наличии. Я пока что, как говорится, здесь, даже вот работаю на кафедре. На этом тогда у нас всё. Всего самого хорошего вам. До свидания.

13. Завершение и вопросы

Благодарности и вопросы

Спасибо, Евгений Вениаминович. Спасибо. Спасибо. Пожалуйста. Спасибо большое. До свидания. Пожалуйста, пожалуйста. Очень рад, что вам что-то там приглянулось.

Остановка записи

Так, теперь надо запись прекратить. Запись кто прекратит? Дмитрий Александрович? Кто прекратит запись? Не прекратил. Наверное, начинала Анна Владимировна, наверное, она должна, да? Закрывать. наверное, или вы? Ну если я нажму, то... Кто включал, тот и выключает, или вы сами? Наверное, нам. Видимо, надо... Сейчас её нет. Ну тогда, видимо, я тогда нажму, да? Да, попробуйте вы сделать это. Только не знаю, что получится. Я не так уж разбираюсь в этой системе. Ну вот так вот. А я вот останавливаю запись.