



Современные методы анализа данных и принятия решений

Ф.Т.Алескеров

Международный центр анализа и выбора решений НИУ ВШЭ

11 ноября 2024

• **Международный Центр Анализа и Выбора Решений**

Научно-учебная лаборатория анализа и выбора решений создана 26 июня 2009 года. 30 июня 2011 г. лаборатория получила статус международной. С 2020 года лаборатория получила статус центра.

Основной целью МЦАВР является проведение исследований в области индивидуального, коллективного и многокритериального выбора, математического моделирования и принятия решений в социально-экономической, политической и финансовой сферах, а также применение теоретико-игровых моделей и методов в экономической теории и анализе данных.

Среди сотрудников МЦАВР 4 ординарных профессора НИУ ВШЭ и 1 Нобелевский лауреат по экономике.

С 2022 года МЦАВР является одним из исполнителей стратегического проекта «Национальный центр научно-технологического и социально-экономического прогнозирования» в рамках программы «Приоритет 2030».

С 2022 года в Международном центре анализа и выбора решений в рамках выполнения стратегического проекта [«Национальный центр научно-технологического и социально-экономического прогнозирования»](#) выполняются исследования в области прогнозирования в условиях глубокой неопределенности.

• Математические модели и практические реализации

1. Сетевые модели

- Анализ и выявление ключевых стран в сетях миграции, торговли, финансовых отношений, международных конфликтов
- Анализ экспорта/импорта нефти, соединений редкоземельных металлов, газа
- Новый способ выявления наиболее цитируемых журналов в сети цитирования

2. Задачи агрегирования

- Найм на работу
- Интегральное качество жизни
- Эффективность функционирования отделений банка, органов исполнительной власти
- Распределение персонала в отделении банка
- Выявление закономерностей выборов в органы государственной власти

3. Паттерн-анализ

- Поиск закономерностей в неоднородных данных
- Типология клиентов в социальных сетях
- Анализ активности в социальных сетях
- Анализ электорального поведения (Выявление закономерностей на выборах в органы государственной власти)

4. Оболочечный анализ данных

- Эффективность мер по лесным пожарам
- Эффективность функционирования университетов
- Филиальная сеть банка

5. Оптимизационные задачи

- Дележ территорий в Арктике
- Вывоз мусора
- Размещение электрозаправок

• Математические модели и практические реализации

6. Моделирование в условиях глубокой неопределенности

- Прогнозирование извержения вулканов, землетрясения, наводнения, разливы нефти, распространения эпидемий с помощью методов сценарного анализа
- Разработка системы поддержки принятия решений для реагирования спасательных служб в условиях ЧС
- Прогнозирование результатов выборов

7. Снижение размерности в данных

- Обработка и анализ больших данных

8. Многокритериальные задачи

- Оценка качества условий проживания
- Анализ деятельности органов исполнительной власти

9. Анализ LC-кривых (временные ряды)

- Анализ покупательской активности клиентов торговой сети, клиентов банка
- Анализ нейросигналов

10. Суперпозиция

- Прогнозирование торнадо
- Поиск в больших данных
- Методы анализа больших данных (получен патент США)

Содержание

1. Продовольственная безопасность
2. Сценарный анализ экспорт/импорта нефти, редкоземельных соединений, газа
3. Развитие зеленой энергетики
4. Разливы нефти в арктических морях
5. Чрезвычайные ситуации природного характера – лесные пожары
6. Чрезвычайные ситуации природного характера – землетрясения, извержения вулканов, наводнения
7. Человеко-машинная процедура выбора природоохранных мероприятий в регионе
8. Модели анализа качества условий проживания населения региона
9. Система оценки качества условий жизни в регионе, городе и муниципалитете
10. Сценарный анализ повышения экономической эффективности территорий на региональном уровне
11. Система размещения курортно-рекреационных комплексов
12. Модели поиска в больших данных

Содержание

13. Сетевые модели конфликтов
14. Сетевые модели взаимодействия террористических групп
15. Новые модели поляризации в сообществах
16. Сценарный анализ миграционных потоков в регионе (группе стран)
17. Механизмы принятия решений в органах расширенного состава БРИКС
18. Модели конкуренции (выборы)
19. Анализ электорального поведения на региональном, районном, городском и муниципальном уровнях
20. Модели сегментации потребительской корзины, планирования кампании по продвижению, оценки потенциала роста продаж, включая адресный прямой маркетинг
21. Снижение экономических потерь, связанных с мошенничеством в сфере страхования
22. Оценка развития филиальной сети коммерческих банков и фирм
23. Оценка значимости коммерческих клиентов для коммерческого банка
24. Модели сегментации жизненного цикла клиента
25. Модель эффективного подбора персонала
26. Библиометрический анализ на основе сетевых моделей

Продовольственная безопасность

Построены сети экспорта-импорта основных зерновых культур (пшеница, рис, кукуруза, просо, сорго, рожь, овес, гречиха, ячмень) между странами на основе реальных данных за 2018-2020 гг.

В построенных сетях определены страны с нехваткой потребления углеводов. Построены различные сценарии изменения экспорта/импорта и производства основных зерновых культур в разных странах.

С помощью новых индексов центральности определены зависимые от импорта основных зерновых культур страны с точки зрения продовольственной безопасности.

Полученные результаты могут быть полезными для улучшения продовольственной политики стран.

Сценарный анализ экспорта/импорта нефти, редкоземельных соединений, газа

Рассматриваются сети экспорта/импорта нефти, редкоземельных соединений, газа между странами.

Для моделирования последствий шоковых событий на рынке рассматриваемых товаров стран используется сценарный анализ.

Построены различные сценарии изменения величин экспорта/импорта нефти, редкоземельных соединений, газа между странами. Определены наиболее зависимые от импорта нефти, редкоземельных соединений и газа страны при возникновении шоковых событий.

Развитие зеленой энергетики

В последние десятилетия проблема изменения климата привела к общемировому пересмотру стратегий развития национальных энергетических систем в пользу более экологически чистых источников энергии. Наиболее активные меры и программы по борьбе с высоким уровнем выбросов парниковых газов были приняты в странах Европейского союза и некоторых соседних с ним государствах. Целью работы является анализ 37 стран Европы за период с 2016 по 2023 гг., включающий в себя выделение групп стран, характеризующихся схожими социально-экономическими показателями (уровень неравенства, подушевой ВВП), структурой выработки электроэнергии, уровнем выбросов CO₂, и определение трендов в структуре энергетических систем.

К возобновляемым источникам относятся солнечная энергетика, геотермальная энергетика, энергия биомассы, гидроэнергетика (русловые, приплотинные, деривационные, плотинные ГЭС), морская энергетика, ветроэнергетика (оншорные и офшорные ветропарки).

Разработанные нами методы туннельной кластеризации и анализа паттернов (в том числе динамического) позволят проанализировать однородность рассматриваемых стран в показателях, отражающих развитие «зелёной» энергетики, а также оценить эффективность мер, направленных на снижение уровня выбросов парниковых газов в последнее десятилетие, и их возможное влияние на экономику и экологическую ситуацию в будущем.

Разливы нефти в арктических морях

Построена универсальная модель, которая может применяться к территории любого масштаба (участку моря, целому морю, всему Северному Ледовитому океану).

Данная модель анализирует рисунок морских течений, ветров и прочих факторов, влияющих на распространение веществ на поверхности моря. Как следствие, упрощается задача по оптимальному расположению сенсоров благодаря меньшему перебору возможных локаций для них.

Предложенная модель является достаточно простой в применении и при необходимости может подстраиваться и усложняться за счёт добавления дополнительных ограничений на этапе оптимизационной задачи или же при изменении правил выбора критических точек.

Чрезвычайные ситуации природного характера

Согласно данным Всемирной метеорологической организации ООН (ВМО ООН), за 50-ти летний период число стихийных бедствий увеличилось в пять раз, однако, благодаря развитию систем прогнозирования, предупреждения и смягчения последствий, количество смертей сократилось почти в три раза. По данным ВМО ООН за период 1970-2019 гг. в мире было зарегистрировано более 11000 стихийных бедствий, в результате которых погибло более 2 миллионов человек, нанесён ущерб около 3,64 триллионов долларов США.

Чрезвычайные ситуации по характеру своего возникновения относятся к полю глубокой неопределённости, что не всегда позволяет заблаговременно и достоверно установить их место, время возникновения и масштаб.

При таких условиях, приоритетным направлением является разработка моделей и методов смягчения последствий ЧС.

Лесные пожары

Разработана модель ранжирования регионов Российской Федерации по степени их уязвимости к лесным пожарам. В основе предлагаемых новых моделей лежит оболочечный анализ данных (Data Envelopment Analysis, DEA). Интервальные методы DEA позволяют работать в ситуациях отсутствия точных данных.

Модель учитывает затраты на охрану окружающей среды, инвестиции в сельское хозяйство и лесоводство, количество пожаров и площадь покрытия пожарами. Модель применяется для ранжирования регионов по степени их уязвимости к лесным пожарам.

Применение интервальных методов DEA дает легко интерпретируемые результаты, открывая большой потенциал для практического применения.

Землетрясения и извержения вулканов

Предлагаемая модель смягчения последствий землетрясений представляет возможность снижения числа жертв и пострадавших на основании многокритериальной системы анализа данных о районах и состоянии зданий и сооружений.

Разработанная имитационная модель извержения вулканов позволяет моделировать различные сценарии действий государственных органов и спасательных служб с учетом временных интервалов, учитывая поражающие факторы вулканической активности, что позволит повысить эффективность ликвидации последствий и спасательных мероприятий.

Наводнения

Модель поддержки принятия управленческих решений построена на основании многокритериальной системы сбора данных о факторах наводнения.

Разработанная модель учитывает класс наводнения, сведения о населении и территории, предусматривает сценарии наводнений и позволяет предлагать государственным органам и спасательным службам наиболее эффективные варианты действий для обеспечения эвакуации населения и расчета данных о пунктах временного размещения.

Человеко-машинная процедура выбора природоохранных мероприятий в регионе

При решении задачи защиты окружающей среды задачи следует учитывать не только существующие условия, но и особенности экономического развития региона. Особенно важно предусмотреть меры, которые уменьшат наиболее серьезное и необратимое воздействие на окружающую среду и негативное влияние на большое количество людей.

Использование стандартных методов не представляется возможным по ряду причин. Так, оценки ситуации для трех основных компонент окружающей среды (воздуха, воды и почвы и растительного покрова) невозможно сравнить в одной шкале.

Цель новых моделей – получить эффективные решения, которые также могут быть легко интерпретируемыми.

Другой целью является предоставление инструментов быстрого поиска среди большого количества возможных мер по защите окружающей среды при разных ограничениях. Эти цели были достигнуты за счет создания человеко-машинной процедуры для выбора природоохранных мероприятий в регионе.

Модели анализа качества условий проживания населения региона



Разработанные модели позволяют выявить проблемные места в развитии региона; определить проблемы, вызывающие наибольшее раздражение населения; построить иерархию проблем по степени их важности и получить рекомендации по их решению; определить наиболее эффективное распределение ресурсов для выполнения проектов, которые позволят максимально повысить удовлетворенность населения. Они позволяют определить и спрогнозировать степень удовлетворенности населения региона в результате предполагаемой реализации проектов (например, строительства дорог, школ, магазинов, реализации экологических проектов и т.п.).

Качество жизни представляется агрегированным показателем, включающим набор показателей, характеризующих инженерную и социальную инфраструктуру, показатели социально-экономического состояния региона.

В разработанной системе используется базовый набор критериев, позволяющих оценивать основные аспекты уровня жизни населения.

Система оценки качества условий жизни в регионе, городе и муниципалитете

Разработана модель оценки качества условий жизни населения в регионе, которая строится на основе анализа основных факторов, описывающих привлекательность территорий для проживания.

Модель оценивает общую комфортность условий в регионе с точки зрения социально-экономических и инфраструктурных характеристик, развития, доступности и качества услуг, предоставляемых региональными подсистемами, уровень обеспечения условий для реализации потенциала развития региона, бизнеса, качества жизни населения.

Модель ориентирована на социальные аспекты состояния и функционирования региона, она учитывает возможности адекватной поддержки устойчивого развития региона в рамках существующих региональных социально-экономических и инфраструктурных подсистем.

Сценарный анализ повышения экономической эффективности территорий на региональном уровне

В модели оценивается территория с точки зрения развития факторов производства для привлечения инвестиций в различных отраслях:

- Развитие промышленности и сельского хозяйства,
- Развитие культурной и коммерческой зоны,
- Развитие индустрии туризма.

Модель направлена на поддержку принятия управленческих решений органов власти на региональном уровне и позволяет повысить экономическую и социальную эффективность и востребованность территорий.

Система размещения курортно-рекреационных комплексов

В процессе разработки генерального плана курортной зоны появляется задача рационального размещения курортных комплексов на территории, принимая во внимание различные экономические, инженерные, социальные и экологические факторы.

Для оценки альтернативных размещений рекреационных объектов формируется набор критериев. Окончательный вариант может быть выбран экспертами из нескольких возможных вариантов, полученных в результате работы модели.

Модели поиска в больших данных

Для решения задачи выбора альтернатив из больших наборов данных разработаны алгоритмы, основанные на суперпозиции надпороговых функций выбора.

Применение разработанных алгоритмов на тестовых наборах данных (например, Microsoft LETOR 4.0) показали высокую (более, чем на 40% выше) эффективность по сравнению со стандартными методами (например, SVM – метод опорных векторов).

Использование метода позволило получить лучшую в мире оценку точности предсказания возникновения торнадо.

Сетевые модели конфликтов

Разработан новый подход к изучению международных конфликтов, основанный на построении сети вовлеченных стран. Принимается во внимание интенсивность международных конфликтов, а также интенсивность внутренних конфликтов в стране. Кроме того, при построении сети учитывается влияние третьих стран на конфликт. Используя предложенные новые индексы центральности в сети, анализируется влияние и участие стран в международных конфликтах.

Рассматриваются три типа конфликта: межгосударственный вооруженный конфликт, внутренний вооруженный конфликт и интернационализированный внутренний вооруженный конфликт.

Для оценки интенсивности конфликта используется информация о количестве погибших в ходе боевых действий, которая была преобразована в ранговые оценки.

Модель применялась к изучению конфликтов в течение ряда периодов времени после Второй мировой войны.

Сетевые модели взаимодействий террористических групп

Произведена оценка центральности действующих лиц в глобальной мировой террористической сети, в которой террористические организации конкурируют между собой за влияние, формируют коалиции.

На основе информации из базы данных Global Terrorism Database (GTD) о террористических атаках в мире за 2001-2018 гг., проведена оценка влияния террористических групп с использованием сетевого подхода и использованием классических и новых индексов центральности.

Предложенные показатели позволяют идентифицировать террористические группы, имеющие прямое или косвенное влияние на террористическую сеть, а также проследить динамику влияния таких групп.

Новые модели поляризации в сообществах

Предлагаются модели распределения мигрантов по районам или регионам страны, направленные на снижение или хотя бы не усиление существующей поляризации населения.

Прибытие новых иммигрантов, представляющих разные культуры и национальности, увеличивает разнообразие принимающего общества. Однако, возникают проблемы неоднородности населения, роста предложения на рынке труда и, самое главное, возможного усиления поляризации между местным населением и мигрантами. Высокий уровень поляризации может привести к напряженности в обществе.

Для достижения оптимального распределения иммигрантов, минимизирующего уровень поляризации в обществе, предлагается несколько моделей, учитывающих разные цели снижения поляризации общества.

Формальная постановка задачи оптимизации включает распределение населения, мигрантов и поляризацию в населенных пунктах.

Сценарный анализ миграционных потоков в регионе (группе стран)

Модель анализирует политическую и экономическую ситуацию, социальную и этнополитическую напряженность в регионе и позволяет предсказать возможную миграцию населения из одного региона в другой.

Используя результаты оценок, правительство может принимать превентивные меры по предотвращению миграции или регулировать его структуру.

Механизмы принятия решений в органах расширенного состава БРИКС

Распределение влияния в выборных органах не прямо зависит от числа голосов участников. Кроме того, результаты коллективного принятия решения зависят от того, как устроены предпочтения участников по формированию коалиций.

Нами предложены новые модели влияния с учетом предпочтений, разработана серия новых индексов, учитывающие желание участников входить в коалиции.

Предложенные индексы более точно измеряют степень влияния в случаях, когда коалиции участников формируются исходя из предпочтений участников о вхождении в коалиции друг с другом.

Модели конкуренции (выборы)

- кандидатов на выборах
- проектов (венчурные проекты);
- распределение ресурсов в сегментированном пространстве (избирательные округа, сегменты рынка, федеральный список).

Анализируется влияние различных факторов на распределение голосов.

Прогноз производится на основании показателей, характеризующих кандидата и конкурентов.

Прогноз распределения голосов представлен в виде интервальных оценок процентов голосов, полученных всеми кандидатами. а также группы показателей, характеризующие социально-экономическую и политическую ситуацию в округе.

Анализ электорального поведения на региональном, районном, городском и муниципальном уровнях

Предложены модели, основанные на том, что избиратели голосуют за партии, в первую очередь, для улучшения своих непосредственных повседневных условий жизни.

Тем не менее, большинство предвыборных кампаний полагаются на опросы общественного мнения, данные которых, несмотря на кажущуюся точность, часто не в состоянии выявить насущные вопросы, позволяющие привлечь избирателей на свою сторону в конкретном избирательном округе.

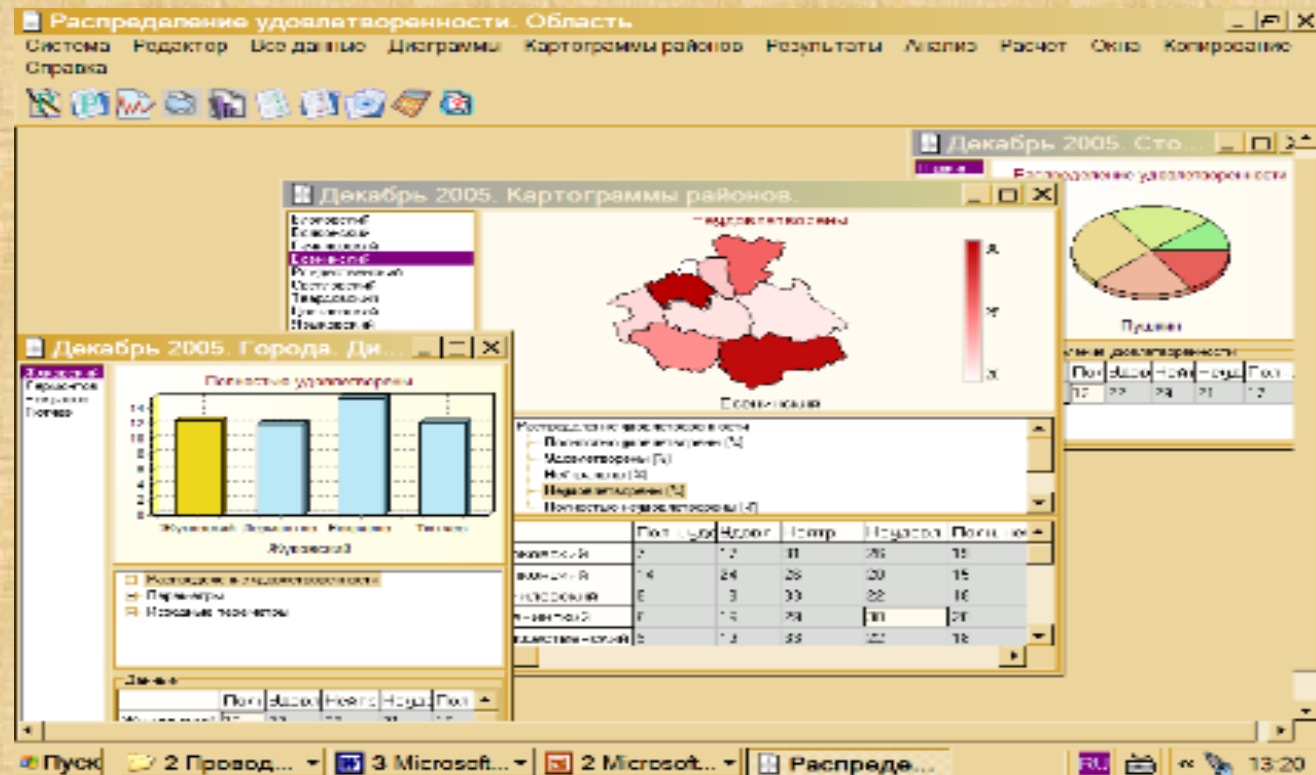
При разработке предложенной модели использовалась эмпирическая информация об условиях, которые влияют на жизнь избирателей непосредственно в местах их проживания. Это дает возможность получить точную, объективную и всестороннюю аналитическую информацию о факторах, определяющих индивидуальный выбор при голосовании.

• Интегральная оценка качества жизни

Заказчик: 11 областей РФ, 17 городов РФ, 2 города за рубежом

Результаты:

1. Сформирована база данных об удовлетворенности населения по регионам России
2. Разработана система поддержки принятия решений позволяющая проводить сценарный анализ развития ситуации, а также предлагать решения для социально-экономического развития регионов



Распределение удовлетворенности по региону

Модели сегментации потребительской корзины, планирования кампании по продвижению, оценки потенциала роста продаж, включая адресный прямой маркетинг

Разработанная модель оценки потенциала увеличения продаж показывает, насколько возможно увеличение продаж для существующих клиентов розничной торговой сети.

Прогноз строится на основе анализа взаимосвязей между товарами и клиентами при совершении покупок. Преимущество этой модели заключается в том, что появляется возможность определить потенциал роста продаж среди существующих клиентов, предлагать персонализированный подход, адресную рекламную и ценовую политику.

Модель была использована фирмой “Metro” для оценки потенциала увеличения продаж в одной из европейских стран и показала высокую эффективность.

Снижение экономических потерь, связанных с мошенничеством в сфере страхования

С начала 2018 года по октябрь 2022 года страховые компании в Российской Федерации направили в правоохранительные органы 39175 заявлений об ущербе, причиненном мошенниками, на общую сумму в 22 миллиарда рублей.

Предложена модель, направленная на выявление факторов, определяющих характеристики групп лиц, участвующих в мошенничестве при получении страхового возмещения на основе сфальсифицированных страховых требований, таких как возмещение в результате дорожно-транспортных происшествий.

Расчеты основаны на семантическом и географическом анализе страховых требований.

Разработанная модель позволит заблаговременно выявлять наиболее подверженные риску отношения в сфере страхования, снизить убытки страховых компаний.

Оценка развития филиальной сети коммерческих банков и фирм

Выбор наиболее перспективных и востребованных населением мест размещения новых филиалов банков и компаний является актуальной задачей повышения экономической эффективности деятельности.

Разработанные модели позволяют производить оценку наиболее перспективных и востребованных мест размещения филиалов, а также производить оценку их эффективности.

Модели также востребованы для оценки экономической эффективности уже существующих филиалов, отделений банков и других предприятий.

Оценка значимости коммерческих клиентов для коммерческого банка

Модель оценивает отношения между коммерческими клиентами и банком на основе существующих данных. В модели определяется относительная эффективность клиентов для банка.

Новые индексы модели разработаны с учетом таких параметров, как прибыльность и оборот. Строятся агрегированные индексы на уровнях типа продукта, категории продукта, которые дают оценку отношений между клиентами и банком.

Новая модель была протестирована на примерах. Сильные стороны разработанной модели заключаются в следующем: модель отражает предпочтения менеджеров, модель проста, не имеет большой вычислительной сложности и позволяет легко выявить причины слабой эффективности.

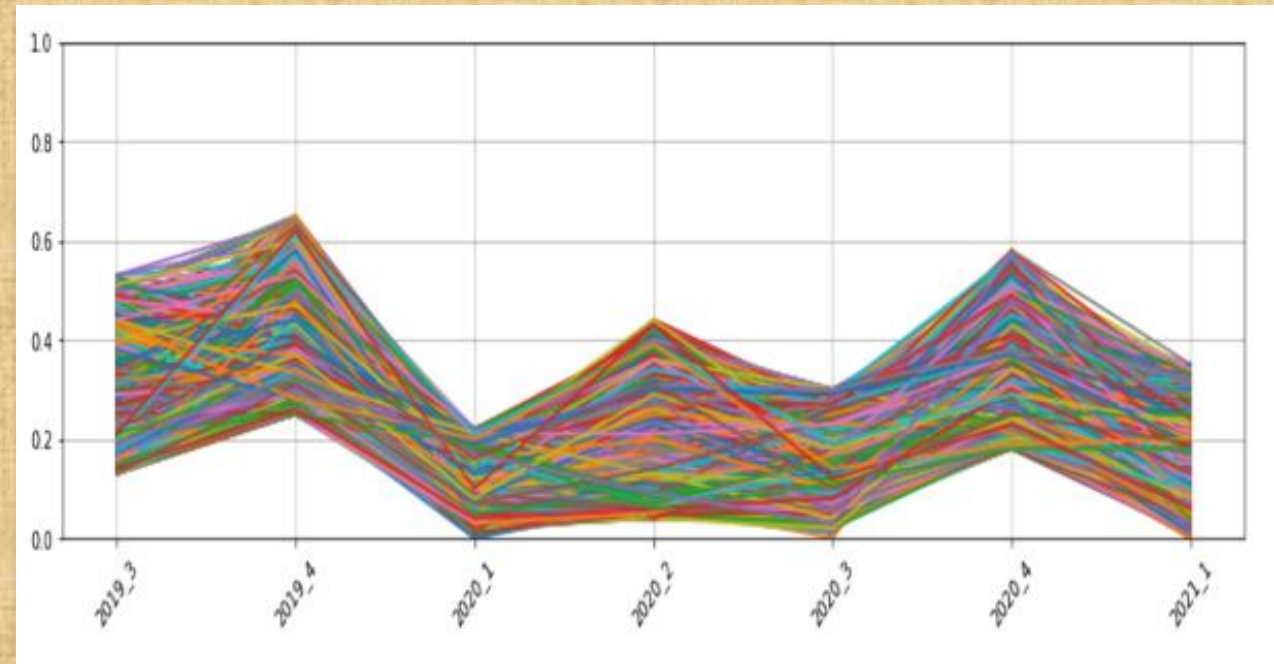
- **Выявление закономерностей в данных транзакций юридических лиц – клиентов Сбер Банка – на основе паттерн-анализа**



Заказчик: Лаборатория искусственного интеллекта Сбера

Результаты:

1. Выявлены закономерности в транзакциях клиентов СберБанка;
2. Проведен поиск «стабильных», «уходящих» и «приходящих» клиентов;
3. Проведена оценка структурных особенностей клиентов и поиск их траекторий развития на основе динамического анализа паттернов;
4. Отобраны макропоказатели регионов для исследования зависимости от транзакций клиентов
5. Сделаны выводы о зависимости макропоказателей регионов от транзакций клиентов



Примеры полученных паттернов на основе более 46 млн. транзакций

Модели сегментации жизненного цикла клиента

Модели разработаны для сегментации существующих клиентов по стадиям жизненного цикла и прогнозирования изменений интенсивности покупок и потребительского поведения.

Типичная область применения:

- Прогнозирование ухода клиентов
- Выработка адресных маркетинговых мер для клиентов на различных этапах жизненного цикла
- Программы увеличения лояльности
- Кредитная политика для клиентов
- Персонализация предложений
- Настройка KPI для персонала в маркетинге и продажах
- Анализ торговой политики в зависимости от продаваемых товаров.

Модель была использована фирмой “Metro” для оценки потенциала увеличения продаж в магазинах фирмы в одной из европейских стран и показала высокую эффективность.

- Сегментация клиентов магазина Metro C&C по данным совершенных покупок

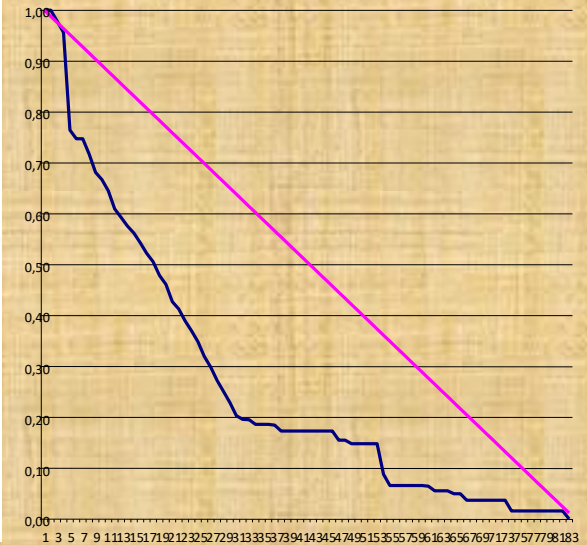
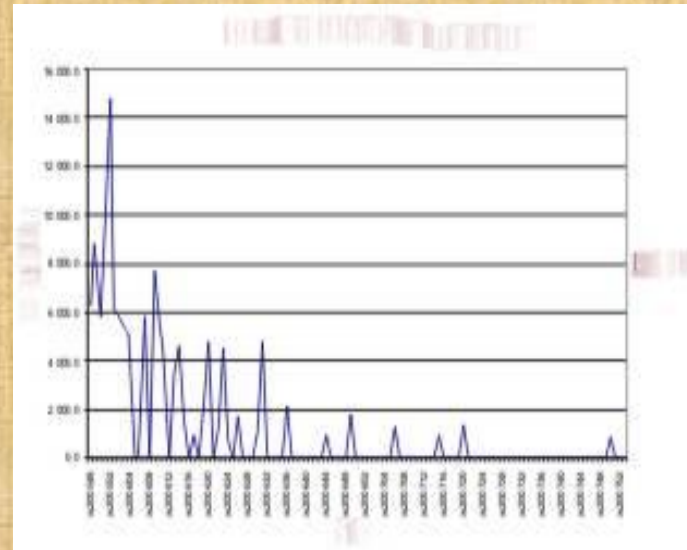


Заказчик: Metro C&C

Результаты:

1. Проведен анализ данных совершенных покупок клиентами Metro C&C
2. По результатам анализа данных о покупках предложена сегментация клиентов Metro C&C
3. Предложены меры по увеличению привлекательности сети магазина для клиентов
4. По результатам данных о сегментации клиентов и предложенным мерам были разработаны программы лояльности для клиентов

Итог: выручка компании Metro C&C увеличилась более чем на 10%



Модель эффективного подбора персонала

Для решения задачи подбора персонала разработана новая математическая модель и создан компьютерный программный комплекс, предназначенный для нахождения эффективных устойчивых соответствий между множествами работодателей и кандидатов.

Учитываются как требования работодателей, так и кандидатов. Комплекс решает задачу автоматизированного подбора персонала. Реализованная модель включает в себя методы сбора данных, преобразование данных и процедуры подбора персонала.

Процедура подбора персонала основана на разработанной математической модели, в которой производится сравнение требований работодателей и кандидатов, применяются различные процедуры агрегирования и строятся соответствия между работодателями и кандидатами.

Библиометрический анализ на основе сетевых моделей

Предлагается комплексный подход анализа публикационной активности с применением методов сетевого анализа. Применяются методы сетевого анализа для исследования паттернов цитирования в, построены сети цитирования для статей, журналов, авторов и их аффилиаций.

Для определения влиятельности того или иного объекта рассчитываются и анализируются классические и новые (предложенные нами) индексы центральности.

В отличие от классических, новые индексы центральности учитывают параметры вершин, дальние взаимодействия между вершинами и групповое влияние вершин друг на друга, что повышает понимание того, как функционируют анализируемые сети.

Спасибо за внимание!