

Краткое резюме некоторых прикладных проектов Рыжова А.П.

1. Системы оценки и мониторинга сложных процессов. Задача оценки и мониторинга включает в себя оценку состояния некоторого процесса и моделирование возможных вариантов его развития на базе всей доступной информации (структурированной, слабоструктурированной, неструктурированной). Такие задачи решаются многими специалистами из различных областей управления экономическими, социальными и политическими процессами. Разработанная технология оценки и мониторинга позволяет делать это максимально эффективно в рамках человеко-машинных систем. К настоящему времени разработаны:

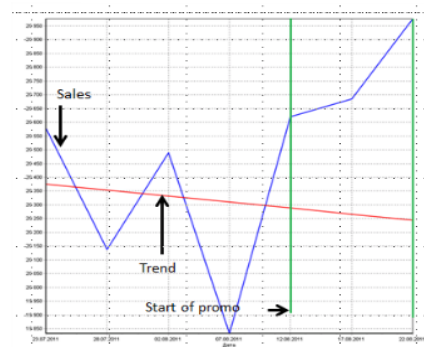
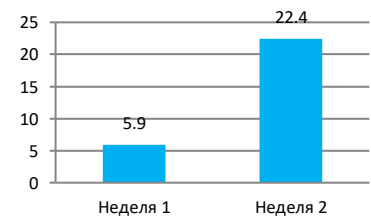
- Система мониторинга и оценки мирной ядерной деятельности стран в интересах управления обеспечения международных гарантий МАГАТЭ (Development of an Intelligent System for Monitoring and Evaluation of Peaceful Nuclear Activity, IAEA, Vienna)
- Система «Мониторинг - 1» (НИЦ «Контур» ФАПСИ РФ)
- Система оценки и мониторинга риска атеросклеротических заболеваний (Государственный научно-исследовательский центр профилактической медицины Минздравсоцразвития РФ)
- Система оценки и мониторинга производства изделий микроэлектроники (Cadence Design Systems, Inc.)

Подробное описание технологии и разработанных систем доступно по [ссылке](#). Разработанное решение имеет широкий спектр приложений и может использоваться в организациях от международного и федерального уровней до предприятий мелкого и среднего бизнеса.

2. Big Data/ Data Mining в телекоммуникационном бизнесе. Бизнес мобильного контента показывает впечатляющую динамику роста на протяжении многих лет. Использование методов интеллектуального анализа данных позволяет значительно повысить его эффективность. Обработка миллионов транзакций и выявление скрытых в них закономерностей поведения клиентов позволяет разрабатывать эффективные рекомендательные системы. Ниже представлены архитектура такого основанного на Big Data/ Data Mining рекомендательного инструмента и результаты его работы.



Рост продаж, %
(7 рекомендованных жанров из 96)



Разработанное решение может использоваться в розничной и интернет-торговле.

3. Big Data/ Data Mining в торговле.

3.1. Профили клиентов.

Вход: база данных чеков + база данных дисконтной программы.

Вопрос: как устроен сегмент высокодоходных клиентов?

Решение: доходность (PROFIT)

разбита на 4 категории (меньше 8200; 8200 – 23300; 23300 – 60500; больше 60500). Строим профиль «PROFIT больше 60500».

Результат: люди, которые часто посещают магазин (в 3,5 раза чаще, чем в среднем по базе), возраста до 30 лет (в 3 раза чаще), покупают мало товаров, средняя сумма чека – маленькая, ...

1	Если	То	Поддержка	Доля	Интерес	Длина
4	if PROFIT is больше 60500 (max 351100)	then => nBUY is больше 100 (max 3800)	25%	87%	3.5	1
6	if PROFIT is больше 60500 (max 351100)	then => AGE is до 30	25%	38%	3.04	1
19	if PROFIT is больше 60500 (max 351100)	then => SUM is меньше 100	25%	68%	2.58	1
23	if PROFIT is больше 60500 (max 351100)	then => nGOODS is [1 - 2]	25%	63%	2.32	1
40	if PROFIT is больше 60500 (max 351100)	then => SEX is 2	25%	75%	1.6	1
104	if PROFIT is больше 60500 (max 351100)	then => TIME is [7-13]	25%	45%	1.25	1
125	if PROFIT is больше 60500 (max 351100)	then => TYPECARD is 10%	25%	100%	1.2	1
223	if PROFIT is больше 60500 (max 351100)	then => DAY is конец месяца	25%	39%	1.12	1
224	if PROFIT is больше 60500 (max 351100)	then => MONTH is апрель	25%	29%	1.12	1
307	if PROFIT is больше 60500 (max 351100)	then => DAY is пн.-чт.	25%	61%	1.08	1
308	if PROFIT is больше 60500 (max 351100)	then => MONTH is весна	25%	57%	1.08	1
309	if PROFIT is больше 60500 (max 351100)	then => MONTH is февраль	25%	26%	1.08	1
309	if PROFIT is больше 60500 (max 351100)	then => MONTH is март	25%	28%	1.05	1
512	if PROFIT is больше 60500 (max 351100)	then => DAY is середина месяца	25%	34%	1.02	1
864						

3.2. Поведение клиентов.

Вход: база данных чеков.

Вопрос: какие товары наиболее хорошо продаются в определенное время (года, недели, дня)?

Решение: разбиваем товары на группы, разбиваем время на периоды, применяем data mining.

Результат:

- для клиентов со средней суммой чека весной характерно приобретать товар группы №10 (в 2 раза чаще случайного поведения);
- для них же зимой характерно совместное приобретение товаров групп № 5, 7, 9;
- ...

2661	if SUM is [1000 - 40000] and MONTH is весна	then => 10	13%	2.11	47%	2	1
2662	if SUM is [1000 - 40000] and MONTH is зима	then => 05 07 09	11%	2.11	63%	2	3
2663	if MONTH is весна and nGOODS is [14 - 105]	then => 13	13%	2.11	54%	2	1
2664	if SEX is 1 and nGOODS is [14 - 105]	then => 01 07	11%	2.11	27%	2	2
2665	if SEX is 2 and SUM is [1000 - 40000]	then => 02 08	11%	2.11	35%	2	2
2666	if SUM is [1000 - 40000] and MONTH is весна	then => 03 09	13%	2.11	31%	2	2
2667	if SUM is [1000 - 40000] and TYPECARD is 10%	then => 10	23%	2.11	47%	2	1
2668	if SUM is [1000 - 40000] and TYPECARD is 10%	then => 05 02 09	23%	2.11	47%	2	3
2669	if SUM is [1000 - 40000] and TYPECARD is 10%	then => 05 07 09	23%	2.11	63%	2	3
2670	if SEX is 2 and SUM is [1000 - 40000]	then => 01 08 09	11%	2.09	42%	2	3
2671	if SUM is [1000 - 40000] and DAY is пн.-чт.	then => 05 03	12%	2.09	30%	2	2
2672	if SUM is [1000 - 40000] and DAY is пн.-чт.	then => 02 08	12%	2.09	35%	2	2
2673	if SUM is [1000 - 40000] and MONTH is весна	then => 05 07 09	13%	2.09	63%	2	3
2674	if SUM is [1000 - 40000] and MONTH is зима	then => 05 02 09	11%	2.09	47%	2	3
2675	if MONTH is зима and nGOODS is [14 - 105]	then => 04	11%	2.09	60%	2	1
2676	if SUM is [1000 - 40000] and DAY is пн.-чт.	then => 05 07 09	12%	2.08	63%	2	3
2677	if SUM is [1000 - 40000] and MONTH is весна	then => 07 08	13%	2.08	48%	2	2
2678	if SUM is [1000 - 40000] and MONTH is зима	then => 10	11%	2.08	47%	2	1
2679	if DAY is пн.-чт. and nGOODS is [14 - 105]	then => 10	13%	2.08	47%	2	1
2680	if SEX is 2 and SUM is [1000 - 40000]	then => 04 07	11%	2.07	42%	2	2
2681	if SUM is [1000 - 40000] and MONTH is весна	then => 02 08	13%	2.07	35%	2	2
2682	if SUM is [1000 - 40000] and MONTH is весна	then => 05 02 09	13%	2.07	46%	2	3
2683	if SEX is 2 and SUM is [1000 - 40000]	then => 05 02 09	11%	2.06	46%	2	3
2684	if SUM is [1000 - 40000] and DAY is пн.-чт.	then => 10	12%	2.06	46%	2	2
2685	if SUM is [1000 - 40000] and DAY is пн.-чт.	then => 03 09	12%	2.06	31%	2	2
2686	if SUM is [1000 - 40000] and DAY is пн.-чт.	then => 02 04	12%	2.06	30%	2	2
2687	if SUM is [1000 - 40000] and DAY is пн.-чт.	then => 05 04 09	12%	2.06	40%	2	3
2688	if SEX is 2 and SUM is [1000 - 40000]	then => 05 03 09	11%	2.05	26%	2	3
2689	if SEX is 2 and SUM is [1000 - 40000]	then => 10	11%	2.04	46%	2	1
2690	if SUM is [1000 - 40000] and DAY is пн.-чт.	then => 06	12%	2.04	46%	2	1
2691	if MONTH is зима and nGOODS is [14 - 105]	then => 13	11%	2.04	52%	2	1
2692	if MONTH is зима and nGOODS is [14 - 105]	then => 10	11%	2.04	46%	2	1

2.3. Совместные покупки.

Вход: база данных чеков

Вопрос: какие товары хорошо продаются вместе?

Решение: разбиваем товары на группы, применяем data mining

Результат:

- группа #14 хорошо продается с товарами групп №5, 4, 7, 9 (почти в 4 раза чаще по сравнению с независимыми покупками);
- добавление товаров групп № 6 или 13 увеличивает продажи;
- ...

2	if GOODS is [14]	then => 05 06 04 07 09	Поддержка	Интерес	Доля	Длина	Длина следствия
2	if GOODS is [14]	then => 05 06 04 07 09	24%	3.64	26%	1	5
3	if GOODS is [14]	then => 05 13 04 07 09	24%	3.62	26%	1	5
4	if GOODS is [14]	then => 05 13 02 07 09	24%	3.59	28%	1	5
5	if GOODS is [14]	then => 05 13 07 08 09	24%	3.59	27%	1	5
6	if GOODS is [14]	then => 05 06 02 07 09	24%	3.58	28%	1	5
7	if GOODS is [14]	then => 05 07 08 10 09	24%	3.57	26%	1	5
8	if GOODS is [14]	then => 05 06 07 08 09	24%	3.56	28%	1	5
9	if GOODS is [14]	then => 05 06 04 07	24%	3.51	27%	1	4
10	if GOODS is [14]	then => 05 13 04 07	24%	3.49	28%	1	4
11	if GOODS is [14]	then => 05 13 07 08	24%	3.49	28%	1	4
12	if GOODS is [14]	then => 06 04 07 09	24%	3.49	28%	1	4
13	if GOODS is [14]	then => 13 04 07 09	24%	3.49	28%	1	4
14	if GOODS is [14]	then => 05 02 04 07 09	24%	3.48	31%	1	5
15	if GOODS is [14]	then => 05 02 04 08	24%	3.47	25%	1	4
16	if GOODS is [14]	then => 05 04 07 08 09	24%	3.47	33%	1	5
17	if GOODS is [14]	then => 06 07 08 09	24%	3.46	30%	1	4
18	if GOODS is [14]	then => 13 07 08 09	24%	3.46	29%	1	4
19	if GOODS is [14]	then => 05 06 07 08	24%	3.45	29%	1	4
20	if GOODS is [14]	then => 05 13 02 07	24%	3.45	30%	1	4
21	if GOODS is [14]	then => 05 13 04 09	24%	3.45	29%	1	4
22	if GOODS is [14]	then => 05 06 04 09	24%	3.44	29%	1	4
23	if GOODS is [14]	then => 05 02 07 10	24%	3.44	27%	1	4
24	if GOODS is [14]	then => 05 07 08 10	24%	3.44	28%	1	4
25	if GOODS is [14]	then => 06 02 07 09	24%	3.43	29%	1	4
26	if GOODS is [14]	then => 05 02 07 08 09	24%	3.43	35%	1	5
27	if GOODS is [14]	then => 05 06 02 07	24%	3.42	29%	1	4
28	if GOODS is [14]	then => 13 02 07 09	24%	3.42	30%	1	4
29	if GOODS is [14]	then => 02 07 10 09	24%	3.42	27%	1	4
30	if GOODS is [14]	then => 07 08 10 09	24%	3.42	28%	1	4
31	if GOODS is [14]	then => 04 07 10	24%	3.39	26%	1	3
32	if GOODS is [14]	then => 05 04 07 08	24%	3.39	34%	1	4
33	if GOODS is [14]	then => 06 13 07	24%	3.38	25%	1	3
34	if GOODS is [14]	then => 05 02 04 07	24%	3.38	33%	1	4
35	if GOODS is [14]	then => 02 04 08 09	24%	3.38	26%	1	4
36	if GOODS is [14]	then => 05 13 08 09	24%	3.36	31%	1	4

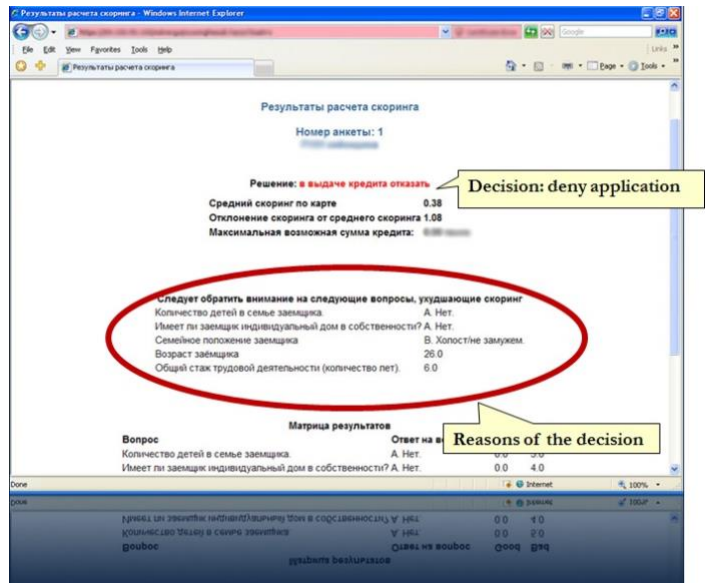
Решения могут быть использованы в торговых организациях от крупных сетей до малого бизнеса.

4. Big Data/ Data Mining в банковском бизнесе: системы кредитного скоринга.

Вход: анкеты + кредитные истории заемщиков. Методика оценки качества кредитов. Технология Data mining.

Выход: интегрированная система автоматической оценки заемщиков.

Результат: в 2 раза меньше дефолтов, в 5 раз меньше «плохих» кредитов.



5. Big Data/ Data Mining в финансах: выявление подозрительных транзакций.

Вход: база транзакций.

Технология: искусственные нейронные сети, нечеткая логика, кластеризация.

Решение:

- Автоматическое выявление нетипичных транзакций.
- Автоматическое выявление транзакций, похожих на указанные пользователем.

Результат: В 7 раз больше «плохих» транзакций по сравнению со случайной выборкой.



6. Big Data/ Data Mining в HR:

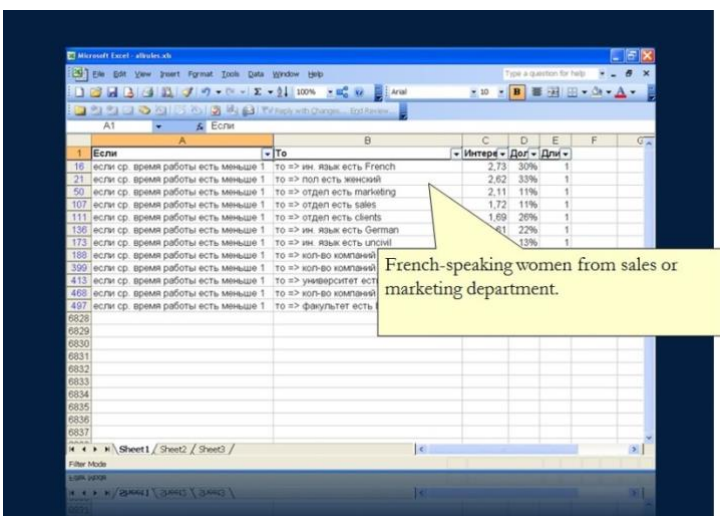
6.1. Анализ резюме, оценка персонала.

Вход: текстовая база резюме.

Вопросы: Кто лоялен, кто нет? Кто растет, а кто стоит на месте? Как собрать эффективную проектную команду? Быстрый отбор кандидатов на должность.

Технология: Ассоциативные правила, нечеткая логика.

Выход: профили лояльных/ эффективных/ и других типов сотрудников, индикаторы лояльности, эффективности, и других качеств (в зависимости от компании).

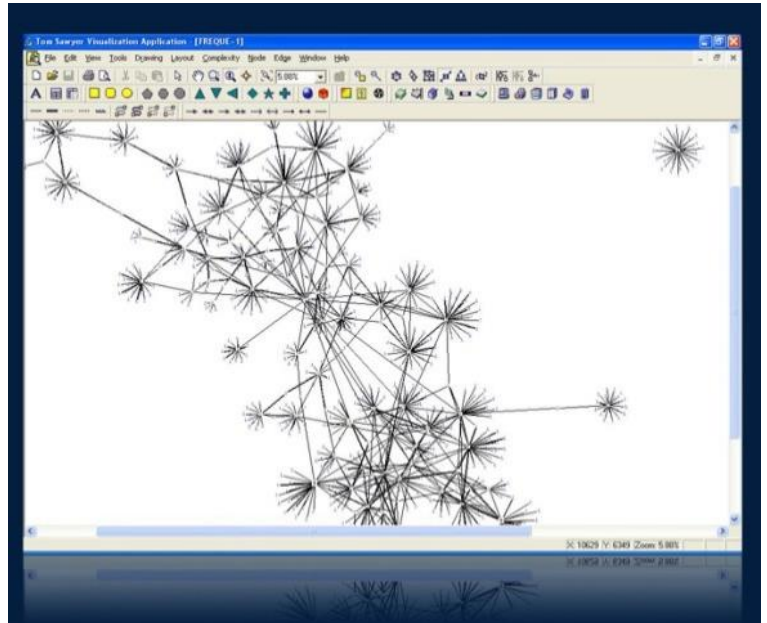


6.2. Анализ и оптимизация коммуникационной структуры компании

Вход: логи корпоративной АТС, exchange server, других используемых ИТ инструментов.

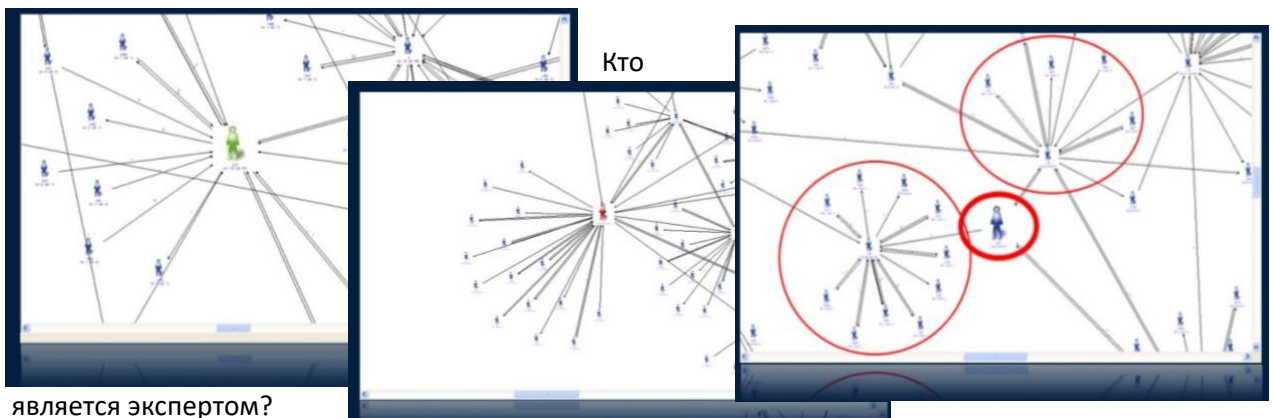
Вопросы: Выявление в компании неформальных группировок и их лидеров:

- Кто является экспертом? К кому обращаются люди?
- Кто является инициатором?
- Кто является связующим звеном между группами?
- Как повысить эффективность коммуникаций и распространение знаний внутри компании/ подразделения/ проектной группы?



Технология: Кластеризация, нечеткая логика, анализ социальных сетей.

Выход: рекомендации по повышению эффективности коммуникаций и распространению знаний



является экспертом?

Как распространяется информация/ знания?

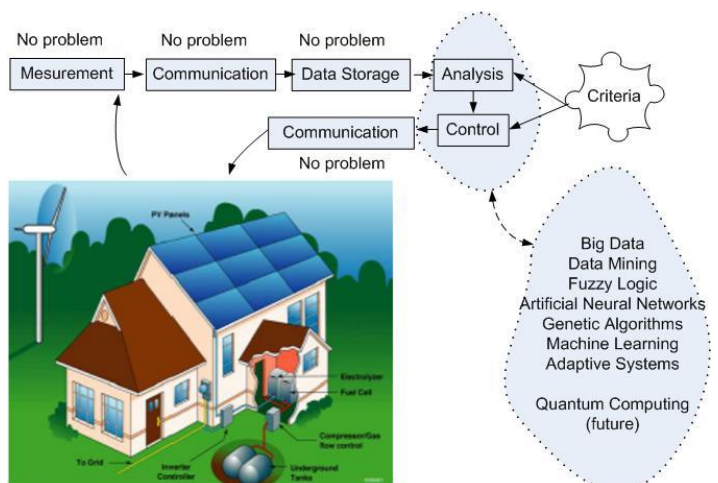
Кто является связующим звеном между группами?

Разработанные решения могут использоваться для оптимизации и повышения эффективности работы организаций крупного, среднего и мелкого бизнеса.

7. Big Data/ Data Mining в энергетике: интеллектуальные сети.

Вход: показания потребления энергии домашними приборами/ домовладениями в рамках подстанции.

Вопросы: Как управлять умным счетчиком/ подстанцией для обеспечения минимизации скачков напряжения (повышения качества электричества)?

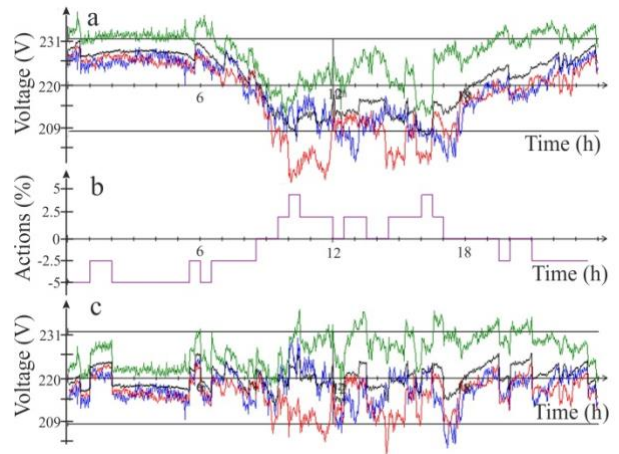


Технология: Ассоциативные правила, нечеткая логика, машинное обучение.

Выход: адаптивный алгоритм управления умным счетчиком/ подстанцией.

Результат: на реальных данных повышение качества электричества в десятки раз.

Решение может быть использовано для оптимизации энергосетей от домохозяйств и дачных участков до городского и регионального уровней.



Суммируя опыт, полученный в ходе выполнения десятков проектов в организациях различного масштаба (от международного и федерального уровней до саппапов) и отраслевой специализации в течение более 25 лет, можно утверждать, что методы интеллектуально анализа данных (Big Data в современной терминологии) являются действенным инструментом повышения эффективности любого бизнеса.

Общая схема использования инструментов Big Data/ Data Mining в бизнесе:

Бизнес – это клиенты

На ком и как мы зарабатываем? На ком и как теряем?

Инструменты Big Data/ Data Mining позволяют обрабатывать транзакции и получать ответы на бизнес-вопросы.

ИТ-инфраструктура позволяет накапливать транзакции.

