

Маркетологи и кризис: как повысить эффективность и профессионализм? Свободное ПО.

Гальченко Максим, руководитель аналитического отдела компании «Холдинг-Центр» (ООО «ЭпиК»),
maxim.galchenko@holding-center.ru

Зачем надо что-то менять?

*Пьем за яростных, за непохожих,
За презревших грошевой уют.
Вьется по ветру веселый Роджер,
Люди Флинта песенку поют.
Павел Коган, Бриганттина*

Как это ни печально, но одними из первых «холодное дуновение кризиса»¹ почувствовали именно маркетологи: количество уволенных коллег уже побило все возможные рекорды. Начинают сокращать расходы именно на отделах маркетинга. Ключевой в данной ситуации является фраза «сокращение расходов». Да, маркетолог-аналитик — сам по себе дорогой специалист, но в придачу к нему необходимо укомплектовать рабочее место — в первую очередь установить программное обеспечение.

Цель статьи как раз и состоит в том, чтобы рассмотреть альтернативные коммерческому ПО пакеты, которые могут существенно снизить затраты на комплектацию такого рабочего места. Более того, используемые в данных пакетах методы и алгоритмы могут существенно обогатить «копилку знаний» и улучшить качество конечных продуктов.

Если рассматривать 1С (в случае анализа продаж) как источник данных для дальнейшей обработки и не увязывать с данной системой «продвинутые» методы анализа, то получится следующая более-менее стандартная комплектация рабочего места:

- MS Office
- SPSS или Statistica
- Дополнительные средства визуализации данных
- Дополнительное специализированное ПО

Оставим за рамками ПО, которое поставляется рядом исследовательских организаций, имеющих сильные IT подразделения, либо же об очень специализированном ПО.

Также лишь вскользь коснемся вопросов, связанных с MS Office, так как только ленивый ничего не слышал об OpenOffice.org.

Поэтому в дальнейшем разговор будет идти о том, что может обеспечить тот функционал, который присущ именно маркетологу-аналитику: анализ данных, визуализация данных.

Вернемся к теме, выделенной в заголовок: зачем менять уже отработанную и понятную структуру? Вот как мне видится ответ на данный вопрос:

- Время, когда над каждым офисом можно было смело вывешивать пиратский флаг безо всякой боязни уже ушло. Первая же проверка лишит хорошо подобранного инструментария, а установивших данное ПО личностей, минимум, спокойного сна и чистой биографии.

«Действовать смело. Никого не расспрашивать. Побольше цинизма. Людям это нравится. Через третьих лиц ничего не предпринимать. Дураков больше нет. Никто для вас не станет таскать брильянты из чужого кармана. Но и без уголовщины. Кодекс мы должны читать.» Ильф и Петров, Двенадцать стульев

- Требовать покупки ПО в ситуации, когда сокращаются расходы, как минимум нелогично.
- Если Вы уволены и фриланс — Ваш хлеб, то всегда может возникнуть вопрос — на чем это все сделано. Кроме того, насколько Вам приятно покупать краденное? А торговать краденным?
- Обычно свободно распространяемое ПО — кроссплатформенное ПО. То есть, Вы абсолютно не привязаны к операционной системе. И Вас абсолютно не заботит, что будет

¹ В зависимости от состояния души заменяем на «последствия нерациональных решений руководства», «последствия странных кадровых решений». Автор не считает «дуновение» настолько уж холодным, но отдает должное сложившимся в последние пол-года традициям. Ну не писать же, «последствия так называемого кризиса, ставшего «козлом отпущения»?

использоваться: Linux или Windows.

- Изучение новых подходов к обработке данных и построению моделей. Вы растете как специалист.

В целом, аргументация сводится к трем принципам: сокращение расходов, независимость от платформы, изучение нового (повышение собственного статуса).

Еще раз акцентирую: темы «А вот SPSS делает это лучше», «В OpenOffice Writer менюшечки другие», «Windows маст дай» мне не кажутся интересными. Такое поведение мне кажется не оптимальным и контрпродуктивным.

«— Голубчики, — сказал Фёдор Симеонович озабоченно, разобравшись в почерках. — Это же проблема Бен Бецалеля. Калиостро же доказал, что она не имеет решения.
— Мы сами знаем, что она не имеет решения, — сказал Хунта, немедленно ощетинаясь. — Мы хотим знать, как её решать.» Аркадий и Борис Стругацкие. Понедельник начинается в субботу

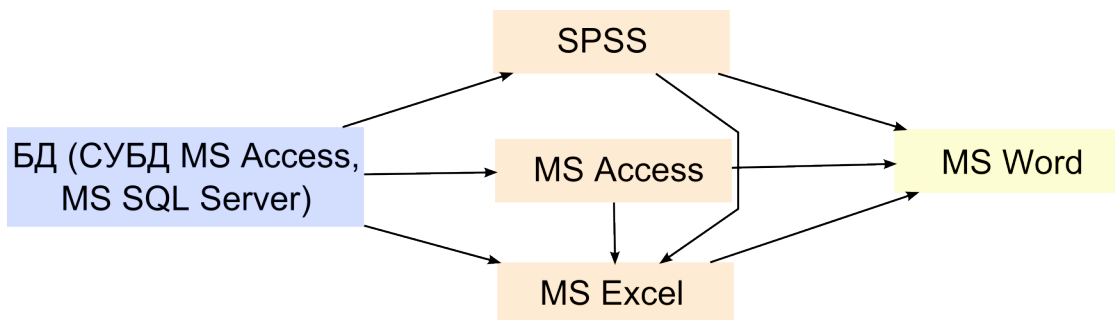
Да, есть в описываемых далее продуктах ошибки и недоработки. Да, придется многому научиться и переучиться. Но это не означает, что свободно распространяемое ПО сплошь ужасно, им ни в коем случае нельзя пользоваться (Вас, скорее всего, кто-то обманул). Все решает баланс плюсов и минусов проприетарного программного обеспечения и свободно распространяемого программного обеспечения. На мой взгляд, на текущий момент плюсов у СПО гораздо больше.

Как дойти до такой жизни?

25 июня. У нас новенький. Китаец. Улыбчивый, усидчивый, прилежный. Приспособили его вместо автоматического потенциометра. Температуру держит идеально. Не отвлекается на перекур, прием пищи.
Июль. Обучали китайца. Теперь активен, курит. В каждом приборе видит предмет многоцелевого назначения. Не верит в каноны и аксиомы. Уже лохмат, не скован традициями и стереотипами. В общем, почти наш человек. Жаль, что скоро уедет. Трудно ему будет одному.
Хроники лаборатории, Юность

А именно — до использования СПО в отделе? Нет ничего легче². Рассмотрим этот процесс на примере компании, в которой я отработал значительное время.

В основном на рабочем месте я всегда занимался анализом данных (статистические данные о производстве, продажах из различных источников). При этом постоянно использовалась следующая схема:



В качестве ремарки - посмотрим на стоимость такого рабочего места, которое де-факто является стандартом в компаниях:

MS Office 2007 Professional Win32 Rus BOX — 22399 руб. (Компьютер-Центр «Кей», апрель 2009 года)

SPSS Statistics Base (SPSS Base) - 44 000 руб. на одного пользователя (официальный сайт SPSS Rus, 1-5 пользователей)

MS Windows Vista Business 32-bit Rus. DVD. OEM — 8 199 руб. (Компьютер-Центр «Кей», апрель 2009 года)

Итого, в данной комплектации имеем: 74 598 рублей. Это с учетом того, что приведена стоимость базовой версии SPSS. В расчете не учитывается стоимость СУБД MS SQL Server, так как данная СУБД обычно используется в корпоративных целях и для решения прочих задач.

Возвращаясь к теме, озвученной в подзаголовке, резюмируем - в зависимости от решаемых задач,

² Исторический срез относится к работе в компании, отличной от текущего местопребывания

соответственно, средством анализа данных у нас были:

- MS Access
- MS Excel
- SPSS

Далее произошло вот что: лицензионная копия MS Word 97 (дело было в далеком 2004 году) отказалась воспринимать созданный ею же файл. Отчет необходимо было сдавать буквально через неделю, но все попытки открыть файл скатывались в «Приложение совершило некорректную задачу и будет закрыто». Вот тогда руководство отдела и решило попробовать применить OpenOffice. О чудо, неродной формат был опознан, файл открыт, а отчет сдан вовремя. Практика оказалась настолько удачной, что OpenOffice прижился в аналитическом отделе. Таким образом, Excel и Word оказались замененными на Writer и Calc.

Сразу хочется отметить несколько особенностей OpenOffice:

- Конечно, возможность сделать экспорт файлов в формат MS Word и MS Excel присутствует, но результат иногда может сильно огорчить. Мы решили данный вопрос путем перехода на PDF формат в общении с внешним миром. А уж вывод в PDF происходит абсолютно замечательно, причем без дополнительной установки конвертеров
- Скорость запуска приложений, входящих в состав OpenOffice может оказаться ниже, чем их аналогов из MS Office
- Что делать, когда нужно продемонстрировать презентацию, допустим? Конвертировать-то нежелательно! Вот для этих целей существует абсолютно замечательная сборка OpenOffice Portable. Тот же самый Impress (аналог Power Point) может быть запущен с флэшки.
- Говорить о полноценной замене MS Access еще очень рано, но OpenOffice Base развивается, и, возможно, скоро мы увидим весьма недурной продукт

И вот, OpenOffice вполне смог заменить MS Word и MS Excel в рассматриваемых схемах.

Далее возник вопрос: а нельзя ли в нашей деятельности полностью отказаться от применения MS Access и такого, безусловно, мощного инструмента как SPSS?

В ходе поиска в Internet возникло несколько интересных вариантов: Tanagra, RapidMiner и KNIME³. Особенно приглянулись два из них: RapidMiner и KNIME. Оба data-mining пакета имеют широкие возможности по интеллектуальному анализу данных, однако, я остановился на KNIME. Причин для этого было несколько:

- Методика разработки пакета такова, что при знании языка программирования Java и среды разработки Eclipse возможно значительное расширение пакета собственными силами, добавление новых возможностей в пакет.
- На экране весьма наглядно предстает сама модель обработки, которая легко конструируется, как в «Лего». Не нужно долго и нудно искать в меню какие фильтры применены, настройки.
- Фактически, нет нужды в написании некоторого скрипта для обработки нового набора данных по утвержденной модели. Подключаем новый набор данных, запускаем на обработку и на выходе — нужные нам результаты.
- Автоматизация процесса построения периодических отчетов
- Огромный выбор алгоритмов, возможность пользоваться мощными средствами языка статистического программирования R.
- Подключение напрямую к БД. Возможность экспорта в Excel, экспорт таблиц в БД.
- Кроссплатформенность. Пакет уверенно чувствует себя как в Linux (во всяком случае, в Ubuntu я не встретил никаких проблем) так и в Windows.

Итак, KNIME⁴. Я опишу данный инструмент чуть более подробно, так как он фактически неизвестен русскоязычному пользователю — в марте месяце количество запросов по слову «knime» в яндексе было равно 14. И то, большую часть набрал ваш покорный слуга.

³ Все три пакета имеют англоязычный интерфейс, но ведь в этом ничего ужасного нет, правда?

⁴ Еще раз хочу акцентировать, я не хочу никого убеждать в том, что KNIME лучше всех, что это лучший выбор, короче говоря, выступать рекламным агентом: бросаем все и переходим. Но я искренне буду рад, если кто-нибудь скачает после прочтения статьи KNIME, попробует и, в лучшем случае, сэкономит некоторую сумму денег.

Обработка данных с помощью KNIME

19 июня. Проходили базовый ликбез. Экскурсия по основным установкам, одну из которых вверяют нам. Мощные гидроприводы. Нагревательная камера. Говорят, рельсы разрывать можно.

23 июня. Действительно, можно. Однако лучше, пользуясь инфракрасными нагревателями, разогревать обеды. Куры в собственном соку тоже получаются ничего.
Хроники лаборатории, Юность

В рамках одного из исследований перед нами стояла следующая задача:

Имеется некоторая сырьевая таблица с данными об объемах продаж светотехнических изделий (порядка 40 000 записей) на территории РФ. В таблице для всех записей определены объемы продаж в денежных единицах. Объем продаж в штуках приведен в приблизительно 70% случаев. Задача: получить полную картину об объемах продаж в штуках, используя классифицирующие признаки: характер товара, страна изготовления, торговая марка, регион, где продавался товар.

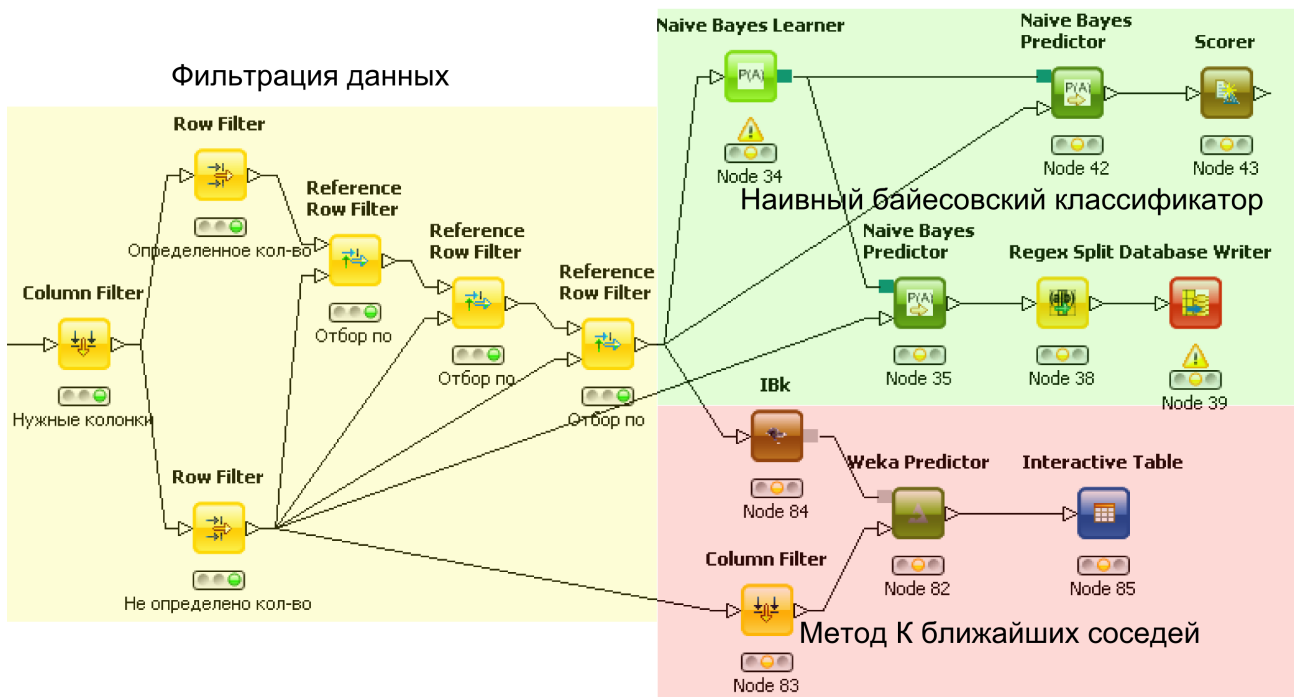
Для решения задачи был применен именно KNIME.

Сначала немного о пакете. KNIME — это разработка University of Konstanz, Германия, кафедры Bioinformatics and Information Mining. Пакет очень непривычен на первый взгляд, так как основан на модели «потока данных».

Собственно, как все это работает: в первом приближении — представьте себе реку. У реки есть источник, на всем протяжении от источника до устья река встречает некоторые препятствия: плотины, островки, река вымывает ил, впадающий в нее ручеек приносит немного каких-нибудь микроэлементов. В результате — вода на выходе значительно изменила скорость течения, изменился ее химический состав.

Аналогия с рекой полная: в качестве воды — данные, островки, ручейки — это узлы (ноды), в которых с данными что-то происходит.

В качестве примера приведу часть потока для описанной выше задачи. Ниже - фрагмент, отвечающий за разделение набора данных на обучающую и целевую выборки, фильтрацию и решение задачи двумя методами - «Наивный Байесовский Классификатор» и «К ближайших соседей». Результат решения с помощью Наивного Байесовского Классификатора возвращается в БД, откуда были взяты исходные данные, в виде таблицы. Подробности того, что делает каждый из узлов скрыты в его настройках, которые можно легко получить выбрав соответствующий пункт контекстного меню. Светофорчик внизу узла показывает выполнен ли он, готов к выполнению или не готов к выполнению.



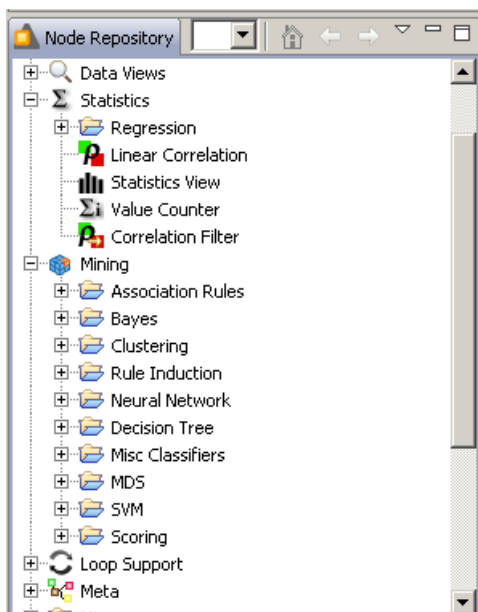
Посмотрите, на схеме все прекрасно видно — сначала отделяются нужные поля (Column Filter), потом поток разбивается на два — поток данных непустым полем «Количеством товара», и поток с пустым полем, далее идет дополнительная фильтрация (все это - желтый сегмент).Затем данные

подаются на построение моделей и прогнозирование двумя методами.

Задача была успешно решена, а результаты пошли в отчет.

Так как над задачей я работал как в офисе, так и дома мне пришлось решать задачу с источником данных — СУБД MS SQL Server в офисе хороша для работы на машинах под управлением Windows, но дома-то Ubuntu. Тут мне помогло умение KNIME спокойно воспринимать СУБД Firebird.

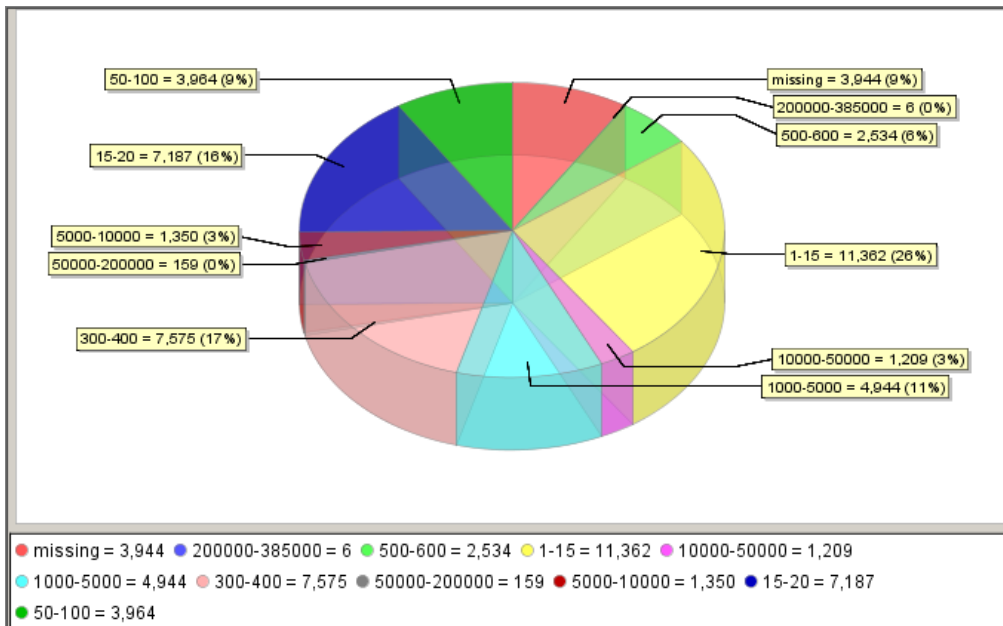
Работа в пакете напоминает конструирование в таком расширенном «Лего». Вы берете кубики (узлы), изменяете их свойства, соединяете друг с другом — и получаете поток данных. На некотором этапе работы с KNIME возникает здоровый исследовательский азарт: а что, если соединить этот узел с этим, а что если использовать другой узел... Такое соотношение взрослых тем и детского желания поиграть с конструктором дает потрясающий эффект — хочется повторять еще и еще, а разнообразие узлов не дает заскучать. Например, оцените количество узлов и групп узлов, с помощью которых можно получить статистику и проводить интеллектуальный анализ данных:



Пакет в базовой комплектации дает возможность:

- Соединяться с БД либо через посредство ODBC, либо с помощью JDBC, либо же считывать данные из файлов с соответствующим форматированием (например, CSV-файлов)
- Преобразовывать исходные данные, фильтровать их, разделять на группы
- Получать базовые статистические характеристики по полям, строить частотные таблицы, корреляционные таблицы, проводить дискретизацию данных

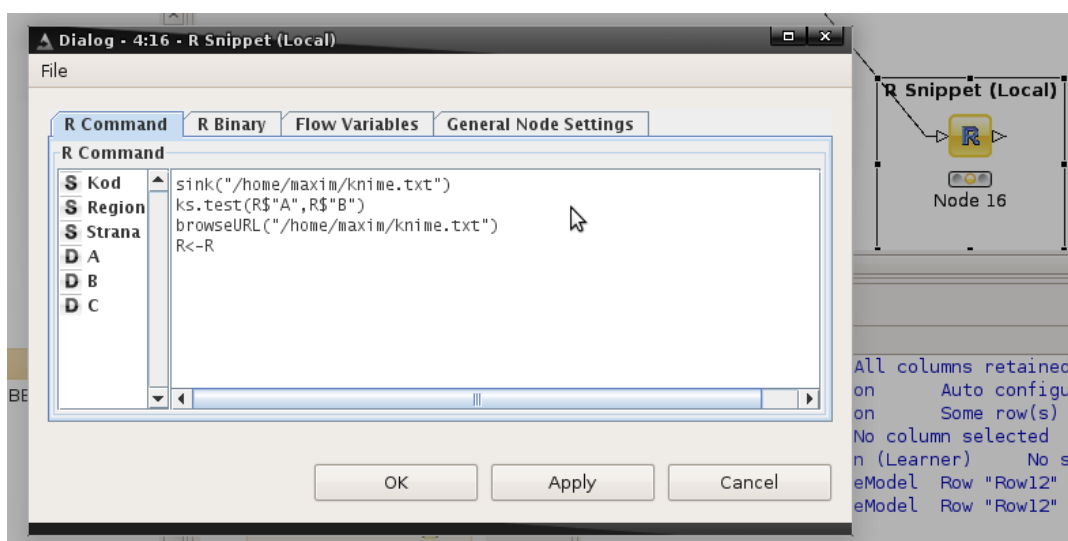
Большой набор узлов непосредственно для интеллектуального анализа данных (data mining) дает возможность эффективно решать все задачи: классификацию, кластеризацию, нахождение скрытых зависимостей в данных. Можно экспортировать таблицы в Excel, есть неплохие возможности для визуализации данных. Вот как, к примеру, может выглядеть круговая диаграмма, сделанная с помощью KNIME.



Базовых функций много, но установка ряда расширений просто необходима. Расширения KNIME устанавливаются просто — либо при первом запуске, либо скачиваются с официального сайта. Сейчас доступно три пакета расширений общего характера (есть еще расширения для химиков):

- WEKA. Данный набор узлов дает доступ к одноименной коллекции алгоритмов data mining. Количество узлов весьма велико, на любой вкус и цвет. Отличное пополнение коллекции Mining.
- R. А вот тут остановимся поподробнее. Когда я начал работать с KNIME, меня несколько насторожило то, что я не вижу знакомых слов: «критерий Колмогорова-Смирнова», «критерий Уилкоксона» и тому подобного. Вот для решения задач, связанных с применением статистических критериев, тестов, в частности, можно использовать данное расширение. Фактически, оно дает некоторый графический интерфейс к языку статистического программирования R. Соответственно, Вы можете реализовать фактически любой сложный расчет, используя возможности данного языка. Если результатом выполнения последовательности операторов является таблица — она может быть передана для дальнейшей обработки в поток KNIME. Отдельный узел предоставляет возможность вывода графиков и диаграмм.

Например, в случае с критерием Колмогорова-Смирнова настройка узла выглядит так:



Вот так, дело ограничивается вставкой трех строк. Вывод идет в текстовый файл. Приведенный выше пример дает в результате следующий вывод:

```
Two-sample Kolmogorov-Smirnov test

data: R$A and R$B
D = 0.2267, p-value < 2.2e-16
alternative hypothesis: two-sided

> proc.time()
пользователь      система      прошло
      1.812          0.088          3.215
```

- Недавно появившееся расширение Textprocessing дает возможность производить интеллектуальный анализ текстов. Эффективность, я думаю, оценят коллеги, которым данное расширение нужно по работе.

Если описывать пакет подробнее — из статьи получится многостраничное руководство пользователя, а это никак не вписывается в формат статьи, но основные возможности, я думаю, мне удалось указать.

Из особенностей можно отметить:

- Низкую скорость запуска и загрузки больших проектов
- Скорость обработки данных в потоке может быть невысокой
- Достаточно большие аппетиты пакета в части потребления оперативной памяти и мощностей процессора

Теоретико-игровые модели

Петрович говорит, что пора начинать работать. Сговорились они, что ли? Обнесли площадку забором и повесили красивые плакаты "Строительство ведет компания "Allstroy" (www.allsteoy.ru)". С чувством выполненного долга играем в Quake. Хроники лаборатории, На стройке

На мой взгляд, использование статистических моделей в ряде случаев не совсем оправдано. Однако, модели именно этого класса чаще всего используются на практике. При этом, в некоторых случаях применение тех же теоретико-игровых моделей может дать гораздо большую информацию о процессах, протекающих на рынке и дать сценарии развития событий в будущем.

Напомню, что теоретико-игровые модели в большей части «завязаны» на описание взаимодействия субъектов, а не статистические данные. То есть, используя знания о принципах взаимодействия на рынке и ограничения Вы можете получить модель оптимального поведения участников рынка. Или оптимальный дележ.

В первую очередь, такая ситуация с применением теоретико-игровых моделей сложилась из-за:

- Сложностей в описании взаимодействия объектов моделирования
- Достаточно серьезной степени проработки статистических моделей и наличию соответствующего ПО
- Сложившемуся стереотипу о том, что те же теоретико-игровые модели — это «игрушки для теоретиков»

Проблема достаточно серьезная и первым шагом на пути к решению является пакет Gambit. Стоит сразу акцентировать внимание на три немаловажных фактора:

- На текущий момент Gambit замер. Активных разработок новой версии не ведется, но вопросы в рассылке задаются и ответы получают
- Gambit может работать только с численными данными — это не пакет символьных вычислений, однако, может воспринимать созданные прочими программными средствами

файлы в собственном формате.

- Поиск всех равновесий по Нэшу может занять значительное время
- Периодически, во всяком случае на ОС Windows, Gambit «взрывает», что выражается в зависании или аварийном завершении работы

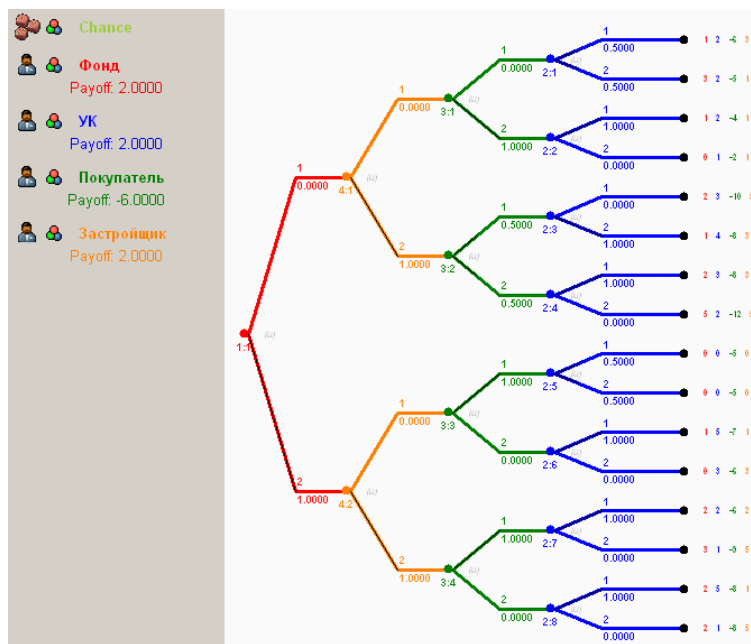
Однако:

- Реализованы несколько методов расчета равновесий по Нэшу, есть возможность преобразования игры в расширенной форме в игру в нормальной форме.
- Есть возможность задать игру как в нормальной, так и в расширенной форме.

То есть, пакет предоставляет возможности по решению игр, но отличается весьма строптивым нравом. И это надо учитывать при попытке работы с ним.

Игра в расширенной и нормальной форме строится достаточно быстро средствами пакета. Интерфейс интуитивно понятен и не вызывает трудностей в работе. Главное, знать и понимать базовые принципы теории игр.

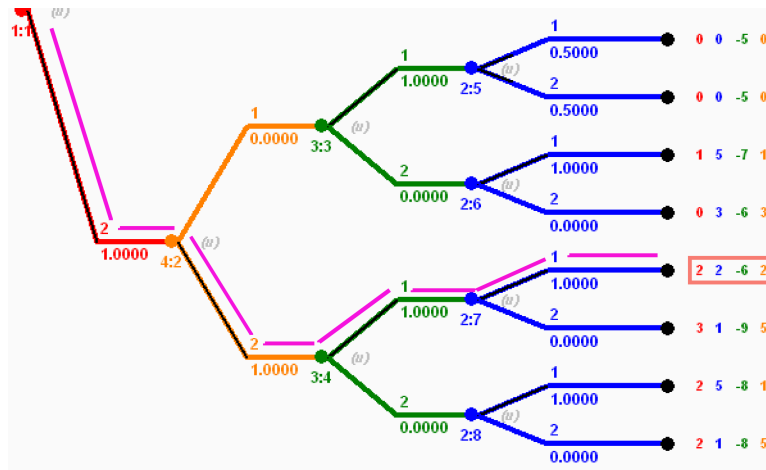
Главное окно Gambit с построенной игрой в расширенной форме выглядит так:



Как видно, играют четыре субъекта. Каждый из игроков имеет два возможных варианта действия в ответ на действие предыдущего игрока. Сначала играет игрок «Фонд», в зависимости от того, какую стратегию выберет «Фонд», далее на ходы «Фонда» и «Застройщика» отвечает «Покупатель», последним принимает решение «УК». Выигрыш «Покупателя» есть сумма выигрышей остальных игроков, взятая с противоположным знаком⁵.

В данной игре все замечательно — существует решение в чистых стратегиях, система замечательно справилась с задачей. Посмотрите на фрагмент дерева, представленный ниже. Система дала ответ: оптимальным будет «Фонду» выбрать вторую стратегию, «Застройщику» ответить «Фонду» второй стратегией, а «Покупателю» и «УК» ответить им первыми стратегиями. Оптимальность подразумевает, что отклонившийся от данной схемы игрок получит меньший выигрыш, если остальные игроки сыграют оптимально. Данный факт легко проверить самостоятельно, «пробежавшись» по остальным ветвям.

⁵ Данная игра была построена исключительно для проверки возможностей пакета.

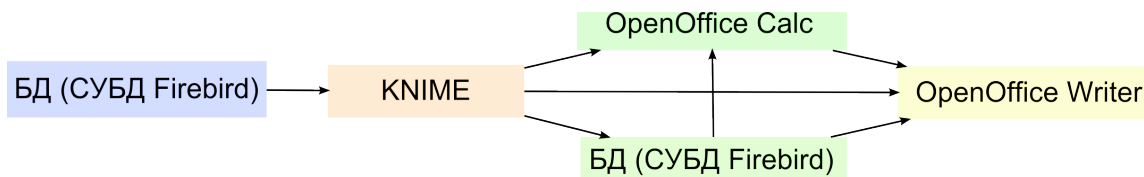


Естественно, в реальных условиях придется серьезно поработать над моделью и тем, как наиболее эффективно применить Gambit. Но, главное, что такая возможность присутствует.

Вместо заключения

*5 декабря. У нас пополнение - новая блондинка с голубыми глазами. Правда, раньше она была парнем в другой лаборатории, где занимались биотехнологиями. По глазам видно - наш человек.
Хроники лаборатории, Звездные войны*

В результате замены проприетарного программного обеспечения на свободно распространяемые аналоги мы построили следующую схему обработки данных:



Стоимость приложений, входящих в данную схему равна нулю. Расходы на внедрение, переход в данной ситуации рассчитываются индивидуально. При этом данный переход может позволить существенно оптимизировать бизнес-процессы.

Для того, чтобы максимально эффективно работать с KNIME, придется освоить на начальном уровне SQL, а также, если есть желание работать с R, и его основы. Однако, потраченное время с лихвой компенсируется возможностями, которые Вы получите.

Работа с теоретико-игровыми моделями даст возможность получать ответы на вопросы, связанные с оптимальным поведением компании на рынке, дележом прибыли. В чисто технических вопросах, связанных с расчетом равновесных ситуаций Вам может помочь Gambit.

Как может отразиться на компании внедрение СПО?

Сильные стороны	Возможности
<ul style="list-style-type: none"> ● Нулевая стоимость закупки ● Получение возможности использовать новейшие технологии ● Поддержка со стороны сообщества продукта ● Возможность участвовать в улучшении продукта, устранении «багов» ● Развитие персонала, повышение квалификации ● Возможность доработать ПО под свои нужды 	<ul style="list-style-type: none"> ● Имидж компании, как «чистой», что важно на репутационных рынках ● Снятие угрозы со стороны «компетентных органов» ● Имидж компании как «высокотехнологичной» ● Предложение новых услуг, основанных на data-mining, text mining, исследовании процессов на базе теоретико-игровых моделей ● Увеличение скорости отклика по периодическим исследованиям

Слабые стороны	Угрозы
<ul style="list-style-type: none"> ● «Перекраивание» бизнес-процессов ● Необходимо переобучить персонал ● Наличие известных и потенциально существующих «багов» (впрочем, от этого не застраховано и коммерческое ПО) ● Возможные проблемы с совместимостью с прочим ПО ● ПО, основанное на Java может быть чувствительным к ресурсам компьютера ● Если ПО первоначально разрабатывалось для ОС семейства Linux, могут возникнуть проблемы с взаимодействием с ОС Windows ● Возможны проблемы с форматами файлов 	<ul style="list-style-type: none"> ● На первоначальном этапе — возможна «лихордка», чем могут воспользоваться конкуренты ● Новый продукт может быть не оценен должным образом на рынке ● Муссирование слухов о ненадежности того, либо иного продукта, раздувание до невероятных размеров обнаруженных проблем

Как видно из приведенной таблицы:

- Часть проблем нивелируются сильным IT-отделом, либо тем, что человек, решивший перейти на данный класс ПО достаточно хорошо обучаем.
- Вообще, данная модель перехода сработает только в том случае, если персонал, либо же Вы, как работник, склонны к познанию нового.
- Необходимо тесное сотрудничество с сообществом и разработчиками продукта, направленное на быстрое исправление ошибок и появлению нового функционала.

Таким образом, переход возможен, а результаты его могут быть весьма благоприятными как для Вас, так и для компании. Удачи!

Полезные ссылки

Официальный сайт KNIME - www.knime.org

Официальный сайт OpenOffice.org - www.openoffice.org

Русскоязычное сообщество OpenOffice.org, адаптированная сборка OpenOffice.org - www.i-rs.ru

Официальный сайт WEKA - www.cs.waikato.ac.nz/ml/weka/

Язык R - www.r-project.org

Gambit - gambit.sourceforge.net