

image not found or type unknown



В последнее время много и много было сказано об анализе информации, которую вы, наконец, можете потерять в проблеме. Хорошо, что многие обращают внимание на такую актуальную тему. Плохо то, что под этим термином все понимают, что им нужно, обычно без общей картины проблемы. Фрагментация в этом подходе является причиной непонимания того, что происходит и что делать. Все это состоит из слабо связанных частей без общего ядра.

Наверняка вы часто слышали выражение «автоматизация пэчворка». Многие уже сталкивались с этой проблемой не раз и могут подтвердить, что главная проблема с этим подходом состоит в том, что почти никогда не возможно видеть общую картину. Ситуация похожа с анализом.

Чтобы понять место и назначение каждого механизма анализа, давайте рассмотрим все это в целом. Он будет основан на том, как человек принимает решения, потому что мы не можем объяснить, как рождается мысль, мы сосредоточимся на том, как мы можем использовать информационные технологии в этом процессе. Первый вариант - лицо, принимающее решения (DM), использует компьютер только как средство извлечения данных и самостоятельно делает выводы. Для решения таких проблем используются системы отчетности, многомерный анализ данных, диаграммы и другие методы визуализации. Второй вариант: программа не только извлекает данные, но и выполняет различные виды предварительной обработки, такие как очистка, сглаживание и т. д. И это применяет математические методы анализа к данным, обработанным таким способом - группировка, сортировка, регрессия и т. д. В этом случае лицо, принимающее решение, не получает необработанные данные, но данные, которые подверглись серьезной обработке, то есть человек уже работает с компьютерными моделями.

В связи с тем, что в первом случае почти все, что связано с самими механизмами принятия решений, возлагается на человека, проблема выбора адекватной модели и выбора методов обработки выходит за рамки механизмов анализа, т. е. Основы для принятия решения - либо инструкция (например, как реализовать механизмы реагирования на отклонения), либо интуиция. В некоторых случаях этого вполне достаточно, но если лица, принимающие решения, заинтересованы, так сказать, в достаточно глубоких знаниях, то просто механизмы извлечения данных здесь не

помогут. Более серьезное лечение необходимо. Это самый второй случай. Все применяемые механизмы предварительной обработки и анализа позволяют лицам, принимающим решения, работать на более высоком уровне. Первый вариант подходит для решения тактических и оперативных задач, а второй - для воспроизведения знаний и решения стратегических задач.

Идеальным примером была бы способность применять оба подхода к анализу. Они могут покрыть почти все потребности организации в анализе деловой информации. Изменяя методы в соответствии с задачами, мы можем в любом случае максимально использовать доступную информацию.

Общая схема работы приведена ниже.

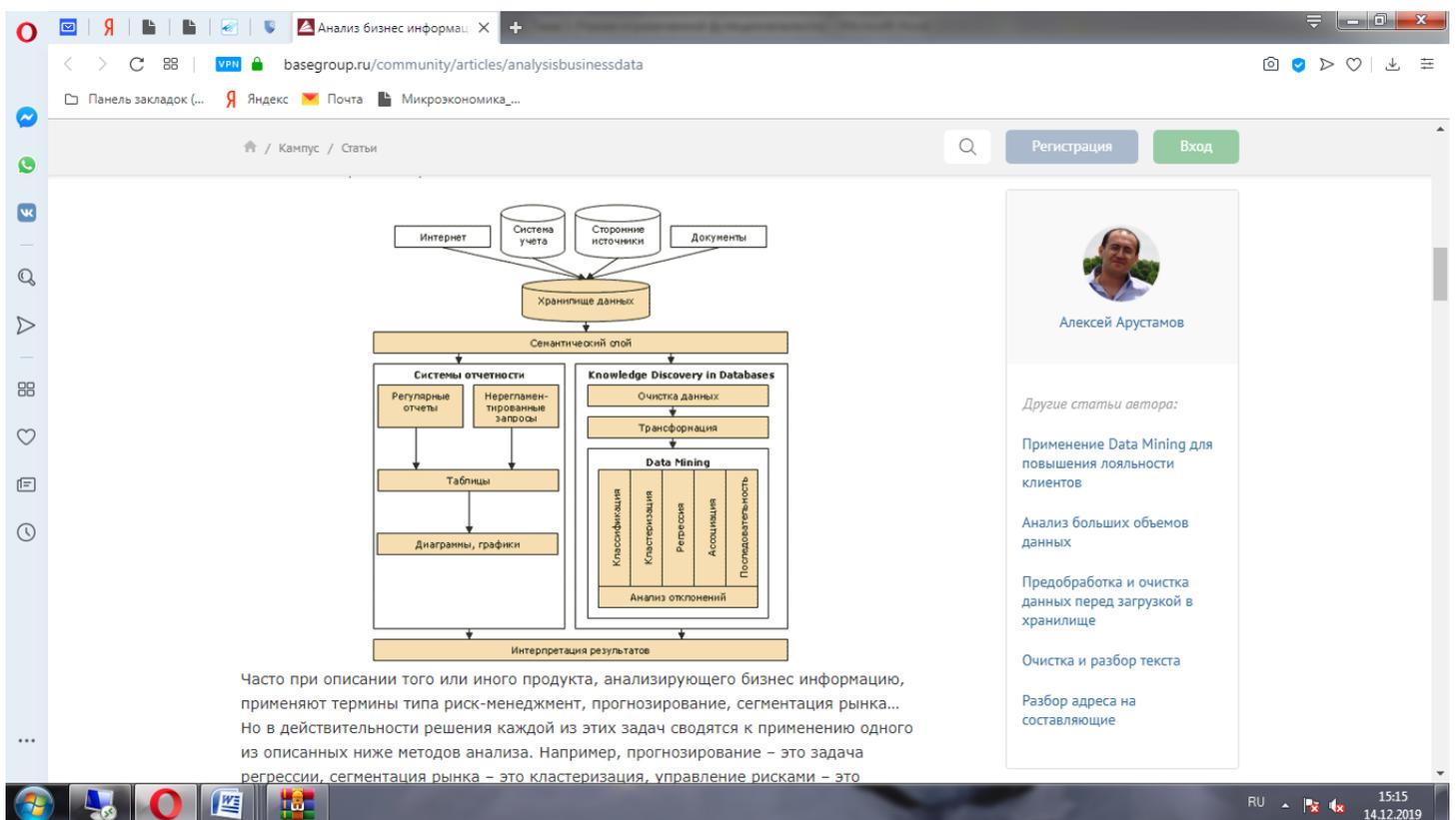


Рисунок 1

Часто при описании продукта, анализирующего бизнес-информацию, используются такие термины, как управление рисками, прогнозирование, сегментация рынка. Но в действительности решения каждой из этих проблем сводятся к использованию одного из методов анализа, описанных ниже. Например, прогнозирование - это регрессионная задача, сегментация рынка - это группировка, управление рисками - это комбинация группировки и классификации, и возможны другие методы. Следовательно, этот набор технологий позволяет решить большинство бизнес-

задач. Фактически, это атомные (базовые) элементы, которые составляют решение конкретной проблемы.

Теперь опишем каждый фрагмент схемы отдельно.

Первичным источником данных должна быть база данных систем управления бизнесом, офисных документов и Интернета, поскольку должна использоваться вся информация, которая может быть полезна для принятия решений. Это не только внутренняя информация для организации, но и внешние данные (макроэкономические показатели, конкурентная среда, демографические данные и т. д.).

Хотя технологии анализа не реализованы в хранилище данных, это является основой для построения аналитической системы. В отсутствие хранилища данных сбор и систематизация информации, необходимой для анализа, займет большую часть времени, что в значительной степени подорвет все преимущества анализа. Действительно, одним из ключевых показателей любой аналитической системы является ее способность быстро получать результаты.

Следующим элементом схемы является семантический слой. Независимо от того, как информация анализируется, необходимо, чтобы она была понятна для лица, принимающего решение, поскольку проанализированные данные в большинстве случаев находятся в разных базах данных, и лицо, принимающее решение, не должно быть погружено в тонкости работы с СУБД проблемы механизмов доступа к базе данных. Семантический слой берет на себя эту задачу. Желательно, чтобы он применялся ко всем аналитическим приложениям, чтобы было проще применять различные подходы к задаче.

Системы отчетности предназначены для ответа на вопрос «что происходит». Первый вариант - использовать его: регулярные отчеты используются для мониторинга рабочей ситуации и анализа отклонений. Например, система ежедневно готовит отчеты о запасе продуктов на складе, и когда его стоимость меньше средней еженедельной продажи, необходимо ответить на это, подготовив заказ на поставку, то есть в большинстве случаев это стандартизированный бизнес операции. Чаще всего некоторые элементы этого подхода в той или иной форме реализуются в компаниях (даже если только на бумаге), однако это не должен быть единственный подход, доступный для анализа данные. Второе применение систем отчетности: обработка специальных запросов. Когда лицо, принимающее решение, хочет проверить мысль (гипотезу), ему нужна пища для размышления,

чтобы подтвердить или опровергнуть идею, потому что эти мысли возникают спонтанно, и нет точного представления о типе необходимой информации, инструмент требуется, позволяя быстро и удобно получать эту информацию. Извлеченные данные обычно представляются либо в виде таблиц, либо в виде графиков и диаграмм, хотя возможны и другие представления.

Хотя для построения систем отчетности можно использовать разные подходы, наиболее распространенным сегодня является механизм OLAP. Основная идея состоит в том, чтобы представить информацию в виде многомерных кубов, в которых оси представляют собой измерения (например, время, продукты, клиенты), а индикаторы (например, объем продаж, средняя цена покупки) размещаются в клетки. Пользователь манипулирует измерениями и получает информацию в нужном контексте.

Благодаря простоте понимания OLAP широко используется в качестве механизма анализа данных, но необходимо понимать, что его возможности в области более глубокого анализа, такого как прогнозирование, чрезвычайно ограничены. Основной проблемой при решении задач прогнозирования является не возможность извлечения представляющих интерес данных в виде таблиц и диаграмм, а построение соответствующей модели. Так что все довольно просто. Новая информация вводится во вход существующей модели, проходит через нее, и в результате получается прогноз. Но построение модели - совершенно нетривиальная задача. Конечно, вы можете поместить несколько готовых и простых моделей в систему, например, линейную регрессию или что-то подобное, довольно часто они делают, но это не решает проблему. Фактически задачи почти всегда выходят за рамки этих простых моделей. И, следовательно, такая модель будет обнаруживать только очевидные взаимосвязи, ценность обнаружения которых незначительна, что уже хорошо известно, или будет давать слишком приблизительные прогнозы, что также совершенно неинтересно. Например, если вы проанализируете рыночную цену на основе простого предположения, что завтра заголовок будет стоить столько же, сколько сегодня, то в 90% случаев вы догадаетесь. И какова ценность этих знаний? Интерес для брокеров составляет всего 10%. В большинстве случаев примитивные модели дают результат примерно на одном уровне.

Правильный подход к построению моделей - это пошаговое усовершенствование. Начиная с первой относительно приблизительной модели, это необходимо, поскольку новые данные накапливаются и модель применяется на практике, уточняйте ее. На самом деле задача создания прогнозов и тому подобного выходит

за рамки структуры систем отчетности, поэтому при применении OLAP не следует ожидать положительных результатов в этом направлении. Для решения проблем более глубокого анализа используется совершенно другой набор технологий в сочетании с названием Database Knowledge Discovery.

Knowledge Discovery in Databases (KDD) – это процесс преобразования данных в знания. KDD включает в себя вопросы подготовки данных, выбора информативных признаков, очистки данных, применения методов Data Mining (DM), пост обработки данных, интерпретации полученных результатов. Data Mining – это процесс обнаружения в "сырых" данных ранее неизвестных, нетривиальных, практически полезных и доступных для интерпретации знаний, необходимых для принятия решений в различных сферах человеческой деятельности.

Привлекательность этого подхода заключается в том, что независимо от предметной области мы используем одни и те же операции:

1. Извлечь данные. В нашем случае это требует семантического слоя.
2. Очистить данные. Запрос на «грязный» анализ данных может полностью свести на нет механизмы анализа, используемые в будущем.
3. Преобразование данных. Различные методы анализа требуют специально подготовленных данных. Например, где-то только цифровая информация может быть использована в качестве входных данных.
4. Собственно выполнить анализ - Data Mining.

Интерпретировать результаты.

Этот процесс повторяется интерактивно.

В свою очередь, Data Mining предоставляет решение только для 6 задач - классификация, группировка, регрессия, ассоциация, анализ последовательности и отклонений.

Это все, что вам нужно сделать, чтобы автоматизировать процесс приобретения знаний. Дальнейшие шаги уже предприняты экспертом, который также принимает решения.

Интерпретация результатов компьютерной обработки возлагается на человека. Просто разные методы дают разную пищу для размышлений. В простейшем случае

это таблицы и диаграммы, а в более сложном - модели и правила. Невозможно полностью исключить участие человека, потому что тот или иной результат не имеет значения, пока он не будет применен к конкретной предметной области. Тем не менее, можно копировать знания. Например, лицо, принимающее решение, используя какой-то метод, определило, какие показатели влияют на кредитоспособность покупателей, и представило это как правило.

Правило может быть введено в систему выдачи кредитов и, следовательно, значительно снизить кредитные риски, поместив их оценку в поток. Кроме того, лицо, занимающееся фактическим декларированием документов, не требует глубокого понимания причин того или иного заключения. На самом деле это перенос методов, когда-то применяемых в отрасли, в область управления знаниями. Основная идея - переход от однозначных и унифицированных методов к транспортным методам.

Все вышеперечисленное, только названия задач. И для решения каждого из них могут применяться различные методы, от классических статистических методов до самообучающихся алгоритмов. Реальные бизнес-проблемы почти всегда решаются одним из вышеперечисленных методов или комбинацией обоих. Практически все задачи - прогнозирование, сегментация рынка, оценка рисков, оценка эффективности рекламных кампаний, оценка конкурентных преимуществ и многие другие - сводятся к описанным выше. Таким образом, имея инструмент, который решает вышеупомянутый список задач, мы можем сказать, что вы готовы решить любую проблему бизнес-анализа.

Список литературы

1. Могилев А. В., Пак Н. И., Хеннер Е. К. Информатика: учеб. пособие для студ. пед. вузов / под ред. Е. К. Хеннера. – М., 2012. – 848 с.
2. Избачков Ю. С., Петров В. Н., Васильев А. А., Телина И. С. Информационные системы: Учебник для вузов. – СПб.: Питер, 2011. – 544 с.
3. Козлов С. В. Применение соответствия Галуа для анализа данных в информационных системах // Траектория науки. – 2016. – Т. 2. № 3 (8). – С. 18.
4. Парватов Н.Г. Соответствие Галуа для замкнутых классов дискретных функций // Прикладная дискретная математика. – 2010. – №2(8). – С. 10-15.
5. Киселева О. М. Пример применения методов математического моделирования // NovalInfo.Ru. – 2016. – Т. 1. – № 42. – С. 67-70.
6. Максимова Н. А. Моделирование образовательной среды личностного развития учащихся // Бюллетень науки и практики. – 2016. – № 5 (6). – С. 481-484.

7. Бояринов Д. А. Новые информационные технологии в системе управления качеством учебного процесса // Известия Смоленского государственного университета. – 2012. – № 4 (20). – С. 464-471.
8. Андреева А. В., Максимова Н. А. ИСУ ВУЗ как инструмент управления качеством образования // В мире научных открытий. – 2013. – № 11.8 (47). – С. 22-28.
9. Баженов Р. И., Лопатин Д. К. О применении современных технологий в разработке интеллектуальных систем // Журнал научных публикаций аспирантов и докторантов. – 2014. – № 3 (93). – С. 263-264.
10. Козлов С. В. Использование соответствия Галуа как инварианта отбора контента при проектировании информационных систем // Современные информационные технологии и ИТ-образование. – 2015. – Т. 2. № 11. – С. 220-225.