

# МЕТОДЫ АНАЛИЗА ЗНАНИЙ В ERP-СИСТЕМАХ

Автор: Коннова Е.П., аспирантка МГТУ им. Н.Э. Баумана  
Научный руководитель: Филиппович Ю.Н., к.т.н., доц. МГТУ им. Н.Э. Баумана

## Необходимость анализа и управления знаниями в ERP-системах

На сегодняшний день системы, автоматизирующие учет деятельности предприятия, оперируют огромными объемами хранящихся оперативных данных. Руководители, управляющие и аналитики предприятий, финансовых учреждений и банков сталкиваются с проблемами оперативного получения аналитической информации, необходимой для принятия стратегических рыночных решений.

ERP представляет собой не класс систем управления, а методологию организации бизнес-процессов предприятия. Важной особенностью проектирования систем управления предприятием на принципах ERP является направленность на планирование ресурсов производства. Большинство функций учета, реализованных в этих системах, служат лишь дополнением к основной задаче — составлению планов поставок материалов, производства и пр. Эволюция функционального развития стандартов от MRP до ERP II представлена на схеме [2]. (Рис.1)

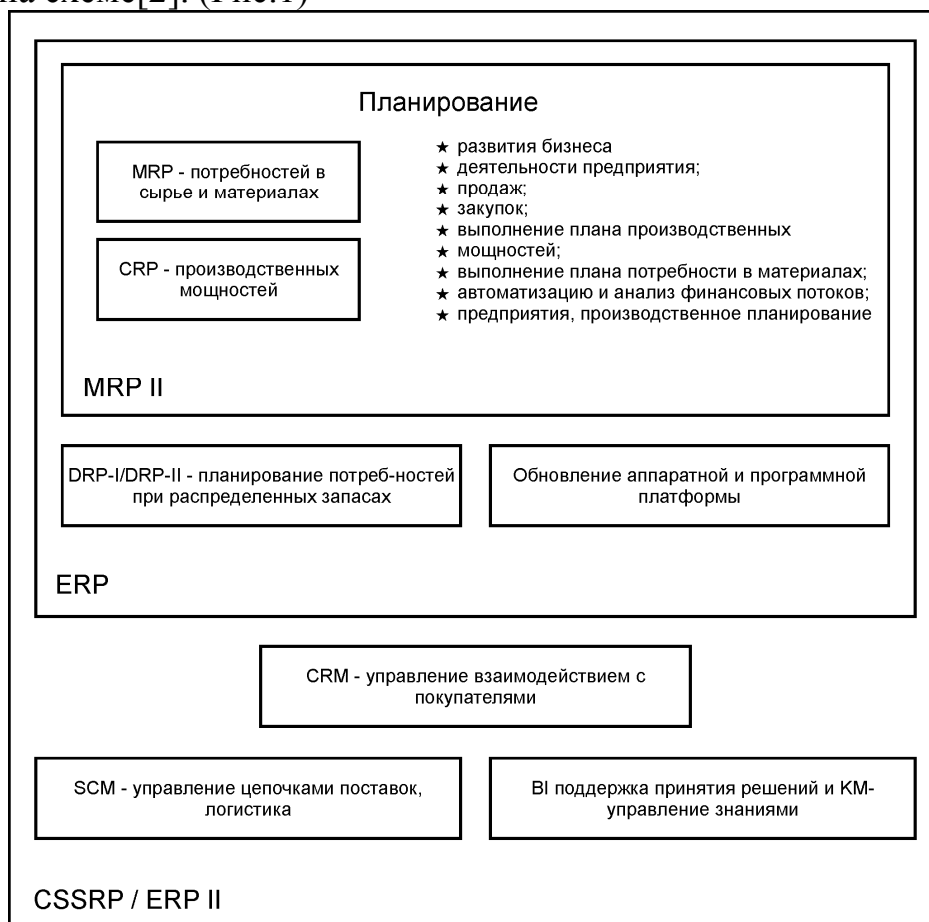


Рисунок 1 Эволюция развития стандарта ERP

Появление новых модулей BI и KM было обусловлено тем[5], что даже успешно функционирующие системы ERP не обеспечивают пользователей систем необходимыми инструментами для работы с огромным объемом хранящихся оперативных данных. Как следствие, стали востребованными возможности систем сбора, управления и анализа бизнес-информации (Business Intelligence, BI) [5].

ERP-системы обладают слабыми возможностями по генерации сложных динамических отчетов с погружением по вертикали и горизонтали отчетов[4] и т.д. Интегрированные аналитические данные обеспечивают предприятия необходимыми типами отчетностей и инструментальными средствами для принятия решений, устраняя ограничения, накладываемые ERP-системой [7]. Например, финансовая информация представлена во многих ERP-продуктах в транзакционной форме, а пользователям нужен быстрый доступ к ней и возможность анализа на различных уровнях детализации. Таким образом, сложились все предпосылки для наращивания ERP-систем инструментами Business Intelligence и поиска методов и алгоритмов для удовлетворения требований и запросов по извлечению и организации знаний в ERP-системах

### **Методы и средства интеллектуального анализа данных**

К категории средств Business Intelligence относят средства представления данных в более удобном для восприятия виде (графики, сводные таблицы, отчеты), позволяющем принимать обоснованные решения[6]. Классификация современных средств Business Intelligence позволяет выделить: генераторы отчетов, средства аналитической обработки данных - клиентские и серверные OLAP- средства, средства поиска закономерностей (Data Mining), средства разработки BI-приложений (BI Platforms) и так называемые Enterprise BI Suites.

Интеллектуальный анализ данных (ИАД, Data Mining) – новое направление в области информационных систем, ориентированное на решение задач поддержки принятия решений на основе количественных и качественных исследований сверхбольших массивов разнородных ретроспективных данных[8]. Представляет собой процесс поиска корреляций, тенденций, взаимосвязей и закономерностей посредством различных математических и статистических алгоритмов. ИАД используется для решения таких задач, как выявление скрытых закономерностей в архивных финансовых, верификация данных о курсах валют, выявление новых потенциальных клиентов, выявление счетов потенциально платежеспособных дебиторов, в различных задачах прогнозирования[8]. Подобные и другие вопросы требуют решения, как на базе предприятий, так и кредитно-финансовых учреждений, в силу пересечения по видам деятельности: работа с клиентами и счетами, анализ финансовой деятельности, принятие решений и т.д.

Все вышеперечисленные достоинства ИАД, доказательства его эффективного применения в силу специфики требований ERP-систем

обуславливают выбор именно этого направления для поиска алгоритмов и методов извлечения знаний. Однако, несмотря на значительный прогресс в области ИАД, практическая реализация этой технологии сопряжена с рядом трудностей. В настоящее время ИАД использует достижения многих разделов современной математики. Многие фирмы-разработчики концентрируют свои усилия на одном-двух конкретных методах, не связывая свои продукты с разработками других фирм в этой области. В результате внедрение средств ИАД существенно усложняется, поскольку многообразие и программных продуктов ИАД, и применяемых в них математических методов затрудняет выбор базового комплекта алгоритмических средств.

## **Методы и средства интеллектуального анализа данных**

Проведем некоторую условную классификацию средств анализа данных, разбив все их множество на четыре направления по методам, реализованным в этих системах.

### ***1. Методы статистической обработки данных.***

1.1 Предварительный анализ природы статистических данных (проверка гипотез стационарности, нормальности, независимости, однородности, оценка вида функции распределения и ее параметров).

1.2 Выявление связей и закономерностей (линейный и нелинейный регрессионный анализ, корреляционный анализ).

1.3 Многомерный статистический анализ (линейный и нелинейный дискриминантный анализ, кластер-анализ, компонентный анализ, факторный анализ).

1.4 Динамические модели и прогноз на основе временных рядов.

Наиболее известные и популярные средства статистического анализа: пакеты Statistica, SPSS, Systat, Statgraphics, SAS, BMDP, TimeLab, Data-Desk, S-Plus, Scenario (BI). Недостатки: требование к специальной подготовке пользователя, большинство методов, основываются на парадигме, в которой главными фигурантами выступают усредненные значения, которые при исследовании реальных сложных жизненных феноменов часто оказываются фиктивными величинами.

Особое направление в спектре аналитических средств ИАД составляют методы, основанные на нечетких множествах. Их применение позволяет ранжировать данные по степени близости к желаемым результатам, осуществлять так называемый нечеткий поиск в базах данных. Однако платой за повышенную универсальность является снижение уровня достоверности и точности получаемых результатов.

### ***2. Кибернетические методы***

***2.1 Методы нейронных сетей.*** Формируются путем построения иерархической сети, узлами которой являются модели нервных клеток

(нейронов), у которых выходной сигнал определяется взвешенной суммой входных сигналов. В свою очередь, входные сигналы представляют собой выходные сигналы нейронов предыдущего уровня. Входными сигналами всей сети являются параметры текущих рядов наблюдений. Ретроспективные данные используются в качестве обучающих выборок, формирующих значения весовых коэффициентов входных параметров нейронов (Рис.2). Основные недостатки: необходимость иметь очень большой объем обучающей выборки; даже натренированная нейронная сеть представляет собой черный ящик. Знания, зафиксированные как веса нескольких сотен межнейронных связей, не поддаются анализу и интерпретации человеком [1](система “KINOSuite-PR”)

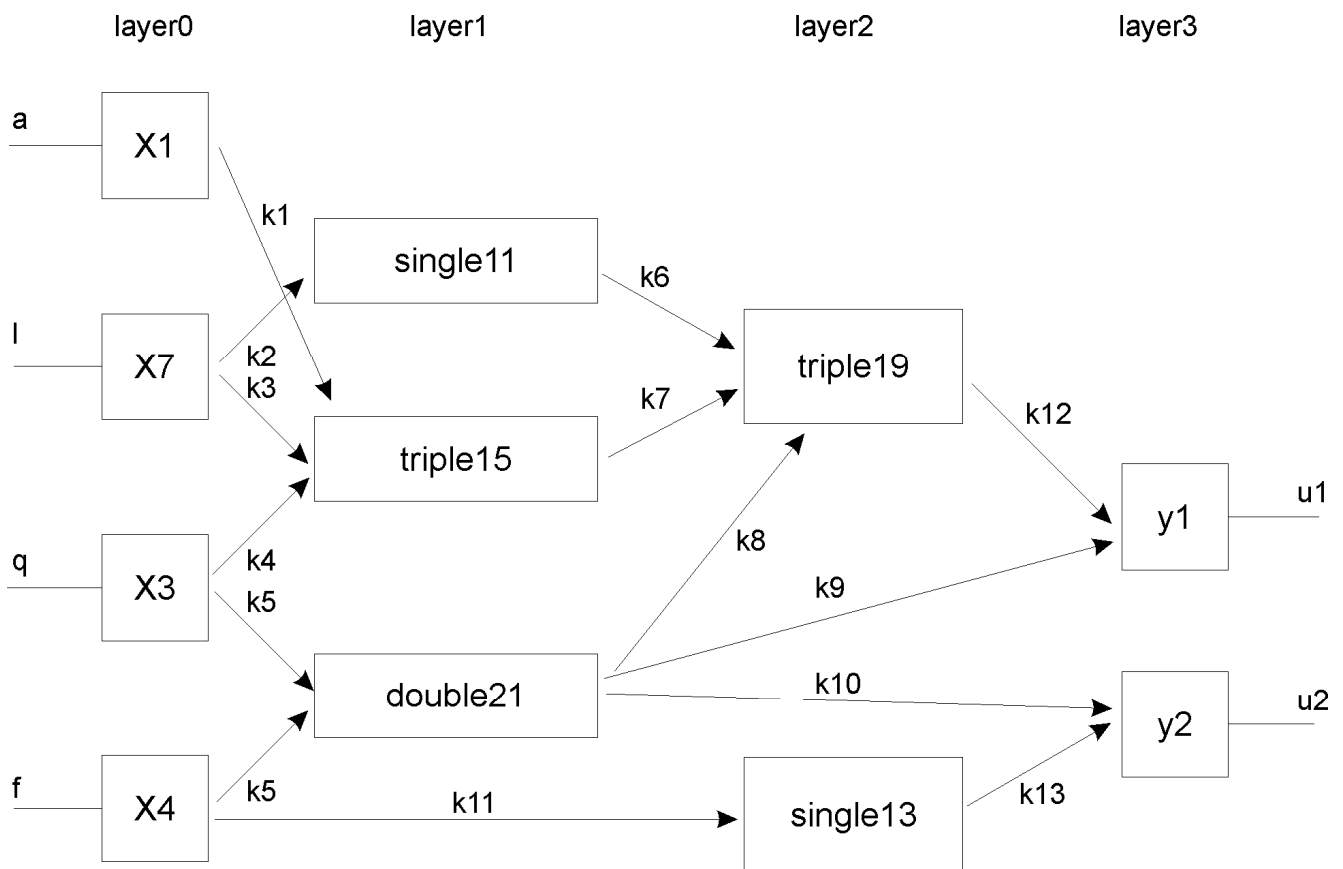
**2.2 Эволюционное программирование.** Получение оптимального решения путем имитации процесса эволюции популяции; вносятся различные, возможно случайные, изменения; совокупность модифицированных решений образует новое поколение возможных решений, которое подвергается «естественному отбору», основанному на «критерии выживания». Процесс итерационно повторяется. Самоорганизующаяся оптимизационная последовательность, приводящая к наилучшему решению (прогнозу) самым неожиданным образом: наиболее эффективное решение может оказаться результатом последовательной эволюции далеко не лучшего (хотя и допустимого) решения. Разновидностью является *метод группового учета аргументов*.

**2.3 Генетические алгоритмы.** Аналогично эволюционному также происходит генерация, отбор и селекция возможных решений (генетических информационных структур с заданным набором параметров («хромосом»)); помимо случайных изменений генной структуры происходит и направленная модификация. Позволяют получать качественно новые, неожиданные результаты решения. Но результаты, полученные кибернетическими методами, часто не допускают наглядных интерпретаций, что усложняет жизнь предметным экспертам (PolyAnalyst(2.2), GeneHunter(2.3), (2.1): NeuroShell, BrainMaker, OWL, 4Thought.).

**2.4 Алгоритмы ограниченного перебора.** Вычисляют частоты комбинаций простых логических событий в подгруппах данных (Бонгард, середина 60-х г.). Ограничением служит длина комбинации простых логических событий. На основании анализа вычисленных частот делается заключение о полезности той или иной комбинации для установления ассоциации в данных, для классификации, прогнозирования и проч. Примеры простых логических событий:

$$X = a; X < a; X > a; a < X < b \text{ и др.},$$

где  $X$  — какой либо параметр, “ $a$ ” и “ $b$ ” — константы. Самым ярким представителем, лидером на рынке продуктов Data Mining этого подхода является система WizWhy. Ее недостаток - выдает решение за приемлемое



$$k7 = 1.5 + 0.4a - 5.3b^2 + 2.3ab + 4.9abc - 0.9c^3$$

**Рисунок 2 Полиномиальная нейросеть**

время только для сравнительно небольшой размерности данных. Тем не менее, система постоянно демонстрирует более высокие показатели при решении практических задач, чем все остальные алгоритмы.

### **3. Традиционные методы решения оптимизационных задач**

Включают в себя вариационные методы, методы исследования операций, включающие в себя различные виды математического программирования (линейное, нелинейное, дискретное, целочисленное), динамическое программирование, принцип максимума Понтрягина, методы теории систем массового обслуживания. Программные реализации большинства этих методов входят в стандартные пакеты: MathCAD и MatLab.

### **4. Экспертные средства**

Средства, связанные с использованием опыта эксперта.

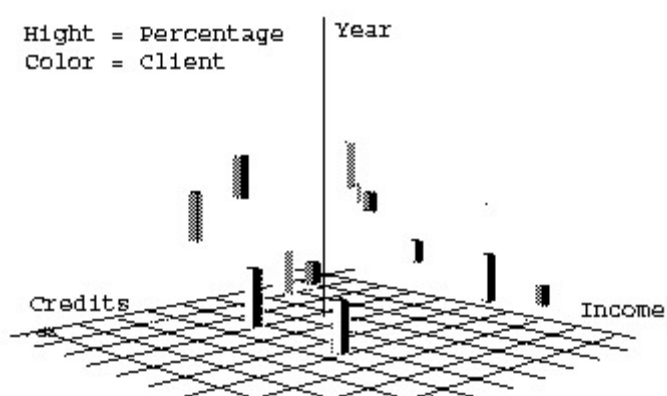
**4.1 Метод «ближайшего соседа»**, лег в основу таких программных продуктов, как Pattern Recognition Workbench или KATE tools. Для осуществления прогноза на будущее или выбора правильного решения в прошлом находятся близкие аналоги наличной ситуации и выбирают тот же

ответ, который был для них правильным. Показывают неплохие результаты в самых разнообразных задачах. Главными их недостатками являются: отсутствие, обобщающих предыдущий опыт, поэтому неизвестно, на основе каких конкретно факторов системы строят свои ответы[3], а также произвол, который допускают системы при выборе меры "близости". От этой меры зависит объем множества прецедентов, которые нужно хранить в памяти для достижения удовлетворительной классификации или прогноза.

**4.2 Деревья решений.** Связан с построением последовательного логического вывода - дерева решений, в каждом узле которого эксперт осуществляет простейший логический выбор («да» - «нет»). В зависимости от принятого выбора, поиск решения продвигается по правой или левой ветви дерева и, в конце концов, приходит к терминальной ветви, отвечающей конкретному окончательному решению[1]. Процесс статистического обучения выведен за пределы программы и сконцентрирован в виде априорного опыта, заключенного в наборе ветвей-решений. (IDIS, See5/C5.0 и SIPINA).

**4.3 Предметно-ориентированные системы.** Это системы анализа ситуаций и прогноза, основанные на фиксированных математических моделях, отвечающих той или иной теоретической концепции. Роль эксперта состоит в выборе наиболее адекватной системы и интерпретации полученного алгоритма[8]. Предельная простота и доступность применения и, как следствие, расплата достоверностью и точностью за эту простоту. (Wall Street Money, MetaStock, SuperCharts, Candlestick Forecaster).

**4.4 Методы визуализации данных.** Нацелены на визуализацию и результатов их анализа, позволяют наглядно отображать полученные выводы для создания у предметных экспертов единой картины ситуации. (DataMiner 3D, Mineset и Impromptu (BI)) (рис. 3)



**Рисунок 3** Пример визуализация данных.

**Зависимость доходов от кредитов под различные проценты и сроки**

## Заключение

Проведение условной классификации по методам, реализованным в этих системах, и анализа средств ИАД позволило прийти к нижеследующим заключениям. Несмотря на обилие методов ИАД, приоритет в эффективных современных разработках смещается в сторону использования моделей представления знаний: логических (дедуктивных и индуктивных), продукционных и графовых (фреймовых). С их помощью решаются задачи прогнозирования, классификации, распознавания образов, извлечения из данных “скрытых” знаний, интерпретации данных, установления ассоциаций в БД. Результаты таких алгоритмов эффективны и легко интерпретируются. Однако известные методы поиска логических правил не поддерживают функцию обобщения найденных правил и функцию поиска оптимальной композиции таких правил. Решением перечисленных проблем можно добиться новых более успешных результатов в области разработок ИАД.

1. Дюк В., Самойленко А.. Data Mining: учебный курс. – СПб: Питер, 2001. –368 с., ил.
2. Когаловский В. Происхождение ERP. Журнал “Директору информационной службы” #05/2000, [www.interface.ru](http://www.interface.ru), 12/03
3. Киселева М., Соломатина Е. Средства добычи знаний в бизнесе и финансах. - Открытые системы, № 4, 1997.
4. Основные проблемы ERP-систем, Журнал JetInfo №2(105)/2002, [www.jetinfo/2002/2/1/article1.2.2002.html](http://www.jetinfo/2002/2/1/article1.2.2002.html), 03/04
5. Программы планирования ресурсов (ERP) необходимы, но недостаточны:.. Март 2000, <http://www.bizcom.ru>, 01/04
6. Федоров А., Елманова Н. Введение в базы данных: средства Business Intelligence, КомпьютерПресс 3'2001, <http://www.olap.ru>, 02/03
7. Шроэк М., Зинн Д., Берг Б.. Интегрированная аналитика. Как извлечь максимальную выгоду из ERP-систем, <http://www.erptomsk.ru>, 11/03
8. Мусаев А. Интеллектуальный анализ данных: Клондайк или Вавилон? Банковские Технологии" , №11-12 1998