

РЕГРЕССИОННЫЙ ПОДХОД К УПРАВЛЕНИЮ СЛАБОСТРУКТУРИРОВАННЫМИ СИСТЕМАМИ

Корноушенко Е.К.

(Институт проблем управления РАН, г. Москва)

ekorno@mail.ru

При решении задач управления слабоструктурированными системами предлагается использовать регрессионный подход. На простейшем гипотетическом примере обсуждаются этапы предлагаемого подхода, условия его применимости, достоинства и ограничения. Показано, как построенные регрессионные модели могут быть использованы для решения статических оптимизационных обратных задач управления в слабоструктурированных системах.

Ключевые слова: слабоструктурированная система, регрессионная модель, линейное программирование, метод наименьших квадратов

Введение

Различным вопросам анализа слабоструктурированных систем уделяется значительное внимание в последнее время. Об этом свидетельствует большое количество работ по данной проблематике, а также рост числа участников ежегодной конференции «Когнитивный анализ и управление развитием ситуаций», проводимой в Институте проблем управления РАН. Возникающие на ней дискуссии касаются, прежде всего, основного вопроса – какой математический аппарат можно считать наиболее приемлемым и перспективным для анализа слабоструктурированных систем, как проверять «правильность» получаемых решений. К сожалению, этот вопрос до сих пор остается открытым.

В данной статье для анализа слабоструктурированных систем предлагается использовать ряд результатов, полученных для

сложных ситуаций в рамках регрессионного анализа. На простейшем гипотетическом примере показано, что если выполняются условия для построения той или иной регрессионной модели, то такая модель может быть использована для решения статических оптимизационных обратных задач управления в слабоструктурированных системах.

1. Современный подход к управлению слабоструктурированными системами

Под *слабоструктурированной системой* (СС-системой) понимается всякая динамическая (т.е. функционирующая во времени) система, в структуре и функционировании которой важную роль играет человеческий фактор. Именно наличие человеческого фактора, для различных проявлений которого практически невозможно построить точные математические модели, позволяет считать такие системы слабоструктурированными системами. «Выразителями» человеческого фактора могут быть отдельные участники системы, структурные подразделения, весь персонал СС-системы и т.д. Поскольку человеческий фактор играет важную роль в функционировании экономических, политических, производственных систем, подобные системы можно рассматривать как СС-системы.

Важность учета человеческого фактора при анализе деятельности компаний наглядно проявилась в подходе, получившем название *сбалансированная система показателей* (*Balanced Scorecard*, см., например, [3,4]). Сбалансированная система показателей (ССП) – это система управления (а не просто система измерения ключевых показателей деятельности компании), которая позволяет компании сформулировать планы на будущее и стратегию и воплотить их в реальные действия. ССП обеспечивает связь между внутренними бизнес-процессами, характеризующими деятельность компании, и внешними измеряемыми показателями, необходимую для повышения стратегической эффективности и достижение намечаемых целей.

По каждому из ключевых аспектов деятельности компании определяются соответствующие наборы возможных действий и измеряемые показатели, оценивающие эффективность этих действий, т.е. задаются цепочки причинно-следственных связей между возможными действиями и результатами этих действий. В плане рассматриваемого ниже регрессионного подхода к управлению СС-системами представляется целесообразным более детально проанализировать некоторые особенности ССП. В деятельности компании выделяются четыре аспекта: финансовый, рыночный, организационный, инновационный. При этом:

- Финансовые показатели, по определению, являются количественными, так что для них не возникает проблем в представлении их стратегических значений и последовательности тактических значений;

- Задание выходных показателей маркетинговой деятельности компании связано с определением интересов клиентов и отражением этих интересов с использованием какой-либо подходящей метрики;

- Показателями хорошей организации внутренних бизнес-процессов могут быть опять-таки соответствующие финансовые показатели (например, себестоимость продукции, рентабельность продаж и т.д.);

- В качестве показателей профессионального уровня персонала могут рассматриваться такие параметры как процент рекламаций на продукцию компании, количество патентов у компании, освоение новых видов продукции и т.п.

Таким образом, при рассмотрении всякой конкретной компании с заданием измеримых выходных показателей особых проблем, как правило, не возникает.

Гораздо серьезнее обстоит дело с определением причинно-следственных цепочек типа «действие – результат», относящихся к каждому из указанных выше аспектов. Дело в том, что внутренние параметры бизнес-процессов могут оказаться настолько взаимосвязаны друг с другом (и с внешними параметрами), что однозначно определяемый результат (в виде наперед заданного изменения рассматриваемого выходного параметра)

будет определяться не столько самим действием, сколько совокупностью условий, при которых это действие совершается. Если проинтерпретировать это обстоятельство в терминах конечно автоматной модели, то получается, что результат совершаемого действия зависит как от самого действия, так и от состояния внутренних процессов на путях, соединяющих точку приложения действия и точку съема информации о последствиях совершения действия, что формально можно представить как

$$Y = \lambda(S, X),$$

где X – вектор значений параметров, характеризующих совершаемое действие,

S – состояние СС-системы, т.е. вектор значений внутренних параметров на указанных путях в структуре системы,

Y – результат действия X .

Сложность ситуации состоит в том, что в силу ненаблюдаемости (и, соответственно, неизмеримости) ряда внутренних параметров СС-системы точное определение состояния СС-системы невозможно и, следовательно, невозможно точное предсказание результата Y . Этот факт существенно затрудняет практическое использование методологии ССП.

Однако если имеется информация о функционировании СС-системы на некотором временном интервале, которую можно рассматривать как совокупность пар (X, Y) , то при определенных допущениях можно построить регрессионную зависимость Y от X , являющуюся некоторым усредненным соответствием

$$(1) \quad Y^* = \lambda^*(S^*, X),$$

для исходной совокупности пар (X, Y) , где S^* – некоторое усредненное состояние СС-системы. Погрешность аппроксимации каждого из соответствий (X, Y) парой (X, Y^*) зависит как от «размаха» (диаметра) самого множества пар (X, Y) , так и от выбора аппроксимирующего соответствия (1). Использование соотношений вида (1), базирующихся на реальных данных о функционировании СС-системы, позволяет ставить и решать (с определенной точностью) некоторые задачи управления СС-системой в формальной постановке.

Цель настоящей работы состоит в том, чтобы:

- не углубляясь в детальный (и, скорее всего, бесперспективный) анализ взаимовлияний внутренних параметров бизнес-процессов друг на друга и на внешние параметры, описать интегральные (усредненные) влияния внутренних параметров на измеряемые внешние параметры в виде тех или иных регрессионных соотношений;

- указать условия и ограничения, при которых справедливы подобные соотношения;

- на простейшем примере показать возможность выбора на основе таких соотношений тех или иных управлений в зависимости от поставленной задачи (здесь рассматриваются лишь статические обратные задачи с оптимизацией заданных критериев).

2. Формирование исходного массива данных

В качестве простейшего примера, на котором будут проиллюстрированы основные этапы предлагаемого подхода, рассмотрим некое гипотетическое предприятие, производящее товары одного функционального назначения, но трех различных типов (скажем, холодильники с разными объемами холодильных камер и разным числом компрессоров).

Допущение 1. Считаем, что все эти товары характеризуются одним и тем же набором признаков (факторов), причем для разных по типам товаров значения (характеристики) некоторых (или всех) признаков могут различаться.

Выделим и систематизируем факторы, которые влияют на указанную прибыль:

- *Производственно-технологические факторы:* вместо перечисления множества детальных факторов, задействованных в процессе производства товаров и определяющих величины постоянных и переменных затрат, используем понятие *себестоимости* (фактор *Себест*) каждого из товаров как результирующий интегральный показатель, характеризующий эту совокупность факторов, и *объем партии* (фактор *Объем*);

- *Организационные факторы:* из совокупности факторов, характеризующих организацию производственных процессов на предприятии выделим два фактора как наиболее отражающих состояние организационной структуры – это *квалификация* (фактор *Квалиф*) обслуживающего персонала (быть может, разная для производства товаров разных типов) а также *неконкурентоспособность* производимых товаров. Квалификация является качественным фактором, и для его измерения используется порядковая шкала (баллы). В качестве количественной меры неконкурентоспособности того или иного вида товара можно рассматривать остатки *непроданных товаров* (фактор *Непрод*) в каждой из выбрасываемых на рынок партий.

- *Рыночные факторы:* это, прежде всего, характеристики самих товаров, совокупность которых определяется в результате анализа спроса на такие товары, объемы проданных на рынке товаров и назначаемые *цены* (фактор *Цена*).

В качестве выходного финансового показателя рассмотрим месячную прибыль предприятия от продажи каждой партии товаров одного и того же типа, произведенных на некотором временном интервале (скажем, в течение 9 месяцев).

Замечание 1. Даже на таком простейшем примере видно, как различные факторы теснейшим образом «переплетаются» друг с другом. Так, квалификация персонала существенно влияет на себестоимость товаров и на их конкурентоспособность, т.е. на число непроданных товаров. Объем партии товара того или иного типа влияет, в принципе, на рыночные цены и на число непроданных единиц этого товара и т.д.

Таким образом, каждая единица товара того или иного типа характеризуется соответствующим вектором значений факторов. По предположению, все эти вектора имеют одинаковую длину.

Допущение 2. Будем считать, что изменения макроэкономических факторов, а также изменения внутри предприятия происходят довольно медленно, в силу чего можно выделить временной интервал T (скажем, год), на котором неизвестная функциональная зависимость между прибылью, получаемой от продажи *каждой единицы товара*, и координатами вектора

факторов сохраняется практически неизменной в течение этого интервала времени.

Это позволяет сформировать исходную выборку в виде массива векторов факторов, описывающих партии товаров, выброшенных на рынок в течение выбранного временного периода, и дополнить этот массив столбцом значений прибыли, полученной в каждый из рассматриваемых моментов времени от продажи каждой из партий товаров. В таблице 1 представлена исходная выборка для предложенного гипотетического примера. Выбранный временной интервал (год) включает 9 моментов времени (3 квартала) поступления рыночной информации о проданных партиях товаров (макроэкономические факторы на интервале T считаются константными и в таблицу 1 не входят).

3. Построение регрессионных соотношений по исходной выборке

Обозначим неизвестную функциональную зависимость, существующую на рассматриваемом интервале T между прибылью, получаемой от продажи каждой партии товара типа k , и координатами вектора факторов, описывающих продаваемые товары, как

$$Y_k = F_k(X_1, X_2, \dots, X_5), k = 1, 2, 3.$$

Для удобства дальнейшего рассмотрения в терминах регрессионного анализа будем называть факторы X_1, X_2, \dots, X_5 – факторами стоимости, а соответствующую прибыль от продажи той или иной партии – зависимой переменной Y .

По предположению, построенная регрессионная модель должна использоваться далее для управления значениями факторов стоимости с целью обеспечения желаемых значений зависимой переменной.

Таблица 1. Исходная выборка: описания товаров в терминах факторов стоимости и результаты продаж

Тип	Себест	Квалиф	Объем	Цена	Непрод	Прибыль
	(тыс.руб.)	(балл)	(ед.)	(тыс.руб.)	(ед.)	(тыс.руб.)
1	8	0.3	260	11	52	208
1	8	0.3	270	11	54	216
1	9	0.3	250	11	45	5
1	9	0.3	230	12	43	174
1	9	0.25	230	12	46	138
1	10	0.25	210	13	40	110
1	10	0.2	210	13	38	136
1	10	0.2	200	14	46	156
1	11	0.1	200	14	42	12
2	20	0.3	175	27	37	226
2	20	0.3	180	27	40	180
2	18	0.3	150	25	36	150
2	18	0.25	150	25	38	100
2	18	0.25	150	26	37	238
2	19	0.2	150	26	32	218
2	20	0.2	130	27	33	19
2	20	0.1	130	27	30	100
2	20	0.1	130	27	32	46
3	35	0.3	64	45	14	10
3	35	0.3	75	45	15	75
3	35	0.5	75	45	13	165
3	33	0.5	75	45	11	405
3	33	0.7	100	46	20	380
3	33	0.7	100	46	17	518
3	31	0.9	150	45	31	705
3	31	1	150	45	28	840
3	31	1	180	44	35	800

Поскольку в любой конкретной регрессионной модели каждый коэффициент отражает характер и силу влияния соответствующего фактора на зависимую переменную, то при использовании такой модели для целей управления зависимой переменной и знаки, и значения её коэффициентов должны не противоречить практическому опыту (и здравому смыслу) разработчика модели (т.е. коэффициенты модели должны быть *предметно интерпретируемыми* – см., например, [1]). Однако, как правило, коэффициенты моделей, полученные с использованием той или иной минимизационной процедуры, не удовлетворяют этим требованиям. Дело в том, что требование минимизации используемого критерия и требование предметной интерпретируемости коэффициентов модели зачастую несовместимы. Это обстоятельство вынуждает переходить к моделям, в которых минимизация выбранного критерия осуществляется *структурно* (т.е. путем последовательного выбора определенной структуры результирующей модели) с сохранением предметной интерпретируемости её коэффициентов.

Ниже предлагается процедура построения одной из таких моделей, базирующейся на значениях коэффициентов корреляции факторов модели с зависимой переменной. Результирующая модель – линейна, а процедура её построения включает следующие этапы:

1) Для каждого из факторов модели находится коэффициент корреляции этого фактора с зависимой переменной;

2) Если этот коэффициент корреляции положителен, то данный фактор входит в модель без изменений. Если же коэффициент корреляции отрицателен, то значения X_i^* данного фактора X_i заменяются на обратные значения вида p_i / X_i , где p_i – порядок наибольших значений фактора X_i (другими словами, вместо фактора X_i рассматривается фактор X_i^{-1} , значения которого определяются указанным выше образом), и находится коэффициент корреляции фактора X_i^{-1} с зависимой переменной Y ;

3) Определяется вклад каждого фактора в среднее значение зависимой переменной Y как произведение среднего значения Y_{cp}

на величину отношения соответствующего коэффициента корреляции к сумме значений коэффициентов корреляции всех факторов модели с зависимой переменной Y ;

4) Определяется среднее значение $X_{i,cp}$ каждого фактора X_i в исходной выборке и вычисляется средняя цена a_i (в единицах измерения зависимой переменной Y) единицы измерения фактора X_i (или фактора X_i^{-1}) как значение отношения вклада этого фактора в величину Y_{cp} к среднему значению $X_{i,cp}$ этого фактора;

5) Результирующая модель имеет вид $Y = \sum_{i=1}^m a_i X_i$.

Вернемся к нашему примеру. Поскольку, по предположению (см. таблицу 1), для каждой партии каждого типа товара известны помесечные прибыли, целесообразно построить модели указанного выше вида для каждого типа товара. Эти модели выглядят следующим образом:

$$Y1 = 32.0063 * \text{Себест}^{-1} + 109.0800 * \text{Квалиф} + 0.0948 * \text{Объем} + 13.3347 * \text{Цена}^{-1} + 0.6420 * \text{Непрод};$$

$$Y2 = -1.9423 * \text{Себест} + 302.0097 * \text{Квалиф} + 0.5546 * \text{Объем} - 0.8294 * \text{Цена} + 1.4594 * \text{Непрод};$$

$$Y3 = 354.7521 * \text{Себест}^{-1} + 163.1473 * \text{Квалиф} + 0.9377 * \text{Объем} + 77.1123 * \text{Цена}^{-1} + 4.4898 * \text{Непрод}.$$

Здесь запись Себест^{-1} (Цена^{-1}) означает, что вместо помесечных данных о себестоимости (цене) в моделях 1 и 3 используются соответствующие данные вида $10/\text{Себест}(i)$, $15/\text{Цена}(i)$, $i = 1, \dots, 9$, где числа 10 и 15 – порядки наибольших значений соответственно себестоимости и цены.

Замечание 2. В модели для товаров 2-го типа факторы Себестоимость и Цена имеют отрицательные коэффициенты. Дело в том, что в силу малости используемой выборки переход к обратным значениям этих факторов не изменил знаки коэффициентов корреляции этих факторов с зависимой переменной. Поэтому в данной модели используются исходные значения себестоимости и цены, а отрицательность вкладов этих факторов интерпретируется как уменьшение прибыли с возрастанием себестоимости (цены) товара.

На рис. 1 и 2 представлены графики месячных прибылей для товаров типа 2 и 3 и модельные оценки этих графиков, полученные с использованием моделей 2 и 3 соответственно.

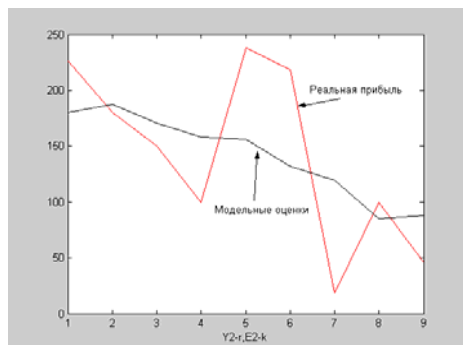


Рис. 1. График месячных прибылей по товарам типа 2 и их модельные оценки

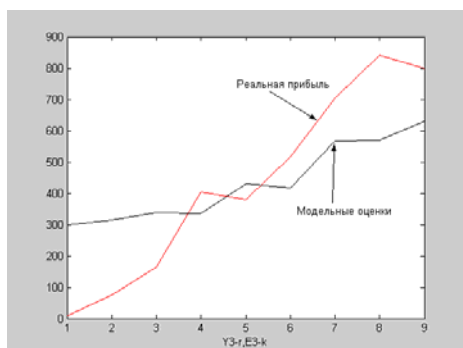


Рис. 2. График месячных прибылей по товарам типа 3 и их модельные оценки

Замечание 3. Качество моделей 1 и 3 (погрешность оценки, коэффициент детерминации и т.п.) не очень хорошее в силу малости исходных выборок по каждому из типов товаров и выбора линейной модели в качестве аппроксимирующей существенно нелинейную зависимость прибыли для товаров каждого

из типов от факторов стоимости. Здесь следует отметить два момента:

- в условиях малой выборки использование более точных нелинейных моделей для аппроксимации исходных данных на обучающей выборке (т.е. на той, по которой строится аппроксимирующая модель) чревато тем, что на других данных, не входящих в обучающую выборку, качество аппроксимации такой модели будет весьма низким (поскольку при её построении весьма значимым может оказаться «эффект малости выборки» — см., например, [2]);
- в плане демонстрации идеи предлагаемого подхода линейные модели весьма наглядны и получаемые результаты легко интерпретируемы.

4. Применение регрессионных моделей для решения некоторых оптимизационных обратных задач управления

Линейная структура построенных регрессионных моделей 1 – 3 позволяет сформулировать и решить некоторые оптимизационные задачи по выбору управлений с использованием линейных алгоритмов минимизации.

Задача 1. Оценить суммарную (по товарам всех типов) максимальную месячную прибыль, которая может быть получена при ограничениях на допустимые значения факторов, приведенных в таблице 2, с наложением дополнительных условий:

- Суммарное (по всем типам) количество товаров, одновременно выбрасываемых на рынок, не превышает 400 единиц;
- При этом суммарное количество непроданных единиц не превышает 70.

Эта задача решается стандартным методом линейного программирования. В частности, её формулировка в вычислительном пакете MATLAB выглядит следующим образом:

$$X = \text{linprog}(F, lb, ub, C, d),$$

где $F=(A1, A2, A3)$, а Ai – вектор коэффициентов i -й модели, lb, ub – интервальные ограничения на допустимые значения факторов, представленные в таблице 2, координаты вектора d суть $d(1)=400, d(2)=70$.

Таблица 2. Нижние и верхние ограничения на допустимые значения факторов стоимости для товаров разных типов

	Себест (тыс.руб.)	Квалиф (балл)	Объем (ед.)	Цена (тыс.руб.)	Непрод (ед.)
Тип 1	12	[0.1,0.3]	[150,300]	[15,18.75]	[30,50]
Тип 2	[15,20]	[0.1,0.4]	[110,200]	27	[25,50]
Тип 3	[20,30]	[0.4,0.8]	[50, 200]	[25, 40]	[10, 40]

Полученное решение, максимизирующее суммарную месячную прибыль, представлено в таблице 3.

Таблица 3. Значения факторов стоимости, являющиеся решением задачи 1

	Себест (т.руб.)	Квалиф (балл)	Объем (единицы)	Цена (тыс.руб.)	Непрод (единицы)
Тип 1	12	0.3	150	15	30
Тип 2	15	0.4	110	27	25
Тип 3	20	0.8	<u>140</u>	25	<u>15</u>

При этом суммарная максимальная месячная прибыль равна 825.7528 (тыс.руб.). Подчеркнутые значения факторов не принадлежат концам допустимых интервалов.

Замечание 4. Малость (непредставительность) исходной выборки не позволяет использовать гипотезу о нормальности распределения месячных прибылей, что делает невозможным оценку доверительного интервала для полученной максимальной прибыли в задаче 1. Это решение следует рассматривать лишь как некоторое среднее значение для возможных значений суммарных месячных прибылей. Следует подчеркнуть, что качество решения задачи 1 в значительной степени определяется

качеством и количеством исходных данных (исходной выборки). Данный пример показывает лишь принципиальную возможность решения указанной задачи в рамках регрессионного подхода.

Задача 2. При ограничениях задачи 1 добиться следующих значений месячной прибыли (в тыс.руб.) по товарам каждого типа:

Вариант 1. Тип 1: 150, Тип 2: 200, Тип 3: 300,

Вариант 2. Тип 1: 200, Тип 2: 300, Тип 3: 400.

Эта задача решается с использованием линейного метода наименьших квадратов с ограничениями. Её формальное представление в пакете MATLAB выглядит следующим образом:

$$X = lsqlin(A,b,lb,ub,C,d),$$

где A – матрица размера 3×15 , на главной диагонали которой в каждой из трех клеток стоят соответствующие вектора $A1, A2, A3$ коэффициентов моделей 1 – 3, а координаты вектора-столбца b суть заданные значения месячных прибылей по товарам каждого типа для вариантов 1 или 2. Остальные параметры lb, ub, C, d определяются так же, как в задаче 1.

Значения месячных прибылей, найденные путем решения задачи 2 для каждого из вариантов, представлены в соответствующем столбце таблицы 4. При этом левые числа являются исходно задаваемыми, а числа после косой линии – соответствующими координатами результирующего вектора решения.

Для полученных решений полностью справедливо замечание 4. Однако даже на этом простейшем примере видно, что требования, приведенные во втором варианте, являются «непопулярными» для предприятия и «в среднем» могут быть выполненными только для товаров типа 3.

Таблица 4. Сравнение исходно задаваемых требований с соответствующими координатами полученных решений

Тип	Месячная прибыль (тыс.руб.)	Месячная прибыль (тыс.руб.)
Тип 1	150/113.3326	200/ 106.1999
Тип 2	200/193.7323	300/ 223.9776
Тип 3	300/ 304.0915	400 / 400

4. Заключение

По представленному материалу можно сделать следующие выводы:

- На простейшем примере показана возможность применения регрессионного подхода для решения некоторых задач управления при анализе СС-систем, в структуре которых содержатся ненаблюдаемые (и неизмеримые) параметры, влияющие на результаты её функционирования;

- Качество результатов анализа и управления СС-системой при использовании регрессионного подхода зависит от представительности исходных данных, по которым строятся используемые регрессионные соотношения;

- Аналогичное рассмотрение может быть проведено и с учетом временных изменений в функционировании СС-системы, что потребует привлечения методов прогнозирования поведения временных рядов. Однако принципиальная возможность применения регрессионного подхода для решения задач управления сохраняется и в этом случае.

Литература

1. АНИСИМОВА И.Н., БАРИНОВ Н.П. ГРИБОВСКИЙ С.В. *Учет разнотипных ценообразующих факторов в многомерных регрессионных моделях оценки недвижимости* // Вопросы оценки, N2, 2004, с. 2–15.
2. БОЛЧ Б, ХУАНЬ К. Дж. *Многомерные статистические методы для экономики*. – М: Статистика, 1979, 263 с.
3. MISSROON A.M. *Demystifying the Balanced Scorecard* // Data Management Review, N12, 1999, pp. 453–467.
4. MISSROON A.M. *Measure vs. Manage* // Data Management Review, N1, 2000, pp. 62–79.