

ПРИМЕНЕНИЕ МЕТОДА DATA ENVELOPMENT ANALYSIS ДЛЯ ОЦЕНКИ ЭФФЕКТИВНОСТИ ИТ-СПЕЦИАЛИСТОВ

Е. П. Моргунов, О. Н. Моргунова

Сибирский государственный университет науки и технологий
имени академика М. Ф. Решетнева

г. Красноярск



«Лучшим каждому кажется то, к чему он имеет охоту» *(К. Прутков)*

Цель доклада — проинформировать уважаемых коллег о методе Data Envelopment Analysis (DEA) и способствовать его продвижению в ваших организациях

1. Понятие эффективности
2. Краткое описание метода DEA
3. Примеры использования метода DEA
4. Программное обеспечение
5. Метод DEA в России
6. Полезные веб-ресурсы



Проблема

Техническая сторона дела:

- базы данных
- разделяемые буфера (shared buffers)
- ядро операционной системы
- и т. д.

А специалисты? Их квалификация и эффективность работы?



Гипотетическая ситуация в компании N

- Проектирование и администрирование баз данных, много различных проектов
- Эти базы данных отличаются друг от друга по количеству таблиц, внешних ключей, столбцов в таблицах, общему объему базы данных (Гб, Тб, число строк) и по другим показателям.
- Работники отличаются по уровню квалификации: высокий и средний
- В течение месяца на каждый проект затрачивается некоторое количество человеко-часов времени специалистов каждого уровня квалификации. Затраты труда – это использованные ресурсы
- Продукцией являются базы данных различной сложности, спроектированные или обслуженные этими специалистами
- Вопрос: насколько эффективно работали наши специалисты в каждом из проектов?

Взгляд на понятие эффективности с двух позиций

- Эффективность — степень достижения цели с учетом затрат ресурсов и времени
 - По-английски — «effectiveness»

- Эффективность =
$$\frac{\text{Результаты}}{\text{Затраты}}$$
 - По-английски — «efficiency»



Эффективность системы

- эффективность — комплексное свойство любой целенаправленной деятельности
- проявляется только в процессе функционирования системы
- отражает степень пригодности системы для ее использования по назначению
- Эффективность системы определяется
 - Используемой технологией функционирования
 - Качеством управления
 - Условиями функционирования
 - Качеством ресурсов
 - Структурой системы



История возникновения метода

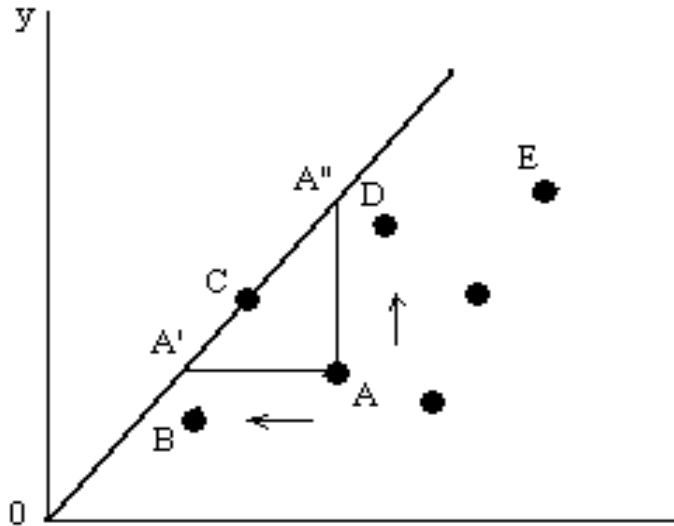
Метод Data Envelopment Analysis (DEA) предложили в 1978 г. американские ученые A. Charnes, W. W. Cooper, E. Rhodes

Charnes, A. Measuring the efficiency of Decision Making Units [Text] / A. Charnes, W. W. Cooper, E. Rhodes // European journal of operational research. – 1978. – Vol. 2. – P. 429–444.

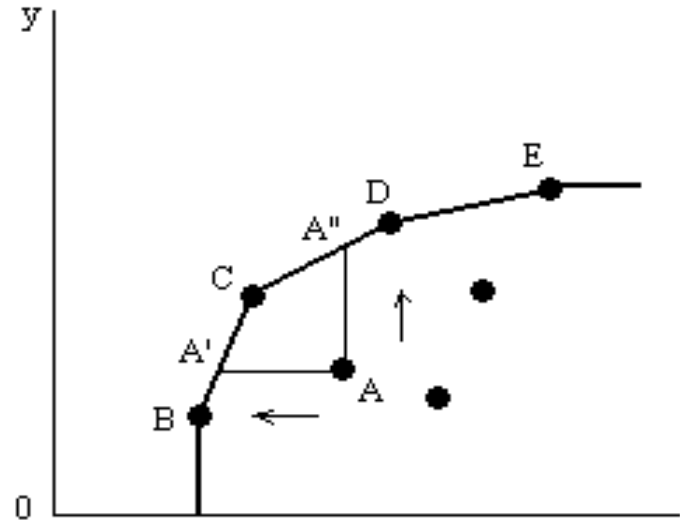
Они основывались на идеях, изложенных в статье M. J. Farrell, опубликованной в 1957 г.

Farrell, M. J. The measurement of productive efficiency [Text] / M. J. Farrell // Journal of The Royal Statistical Society, Series A (General), Part III. – 1957. – Vol. 120. – P. 253–281.

Идея метода DEA



Постоянный эффект масштаба



Переменный эффект масштаба

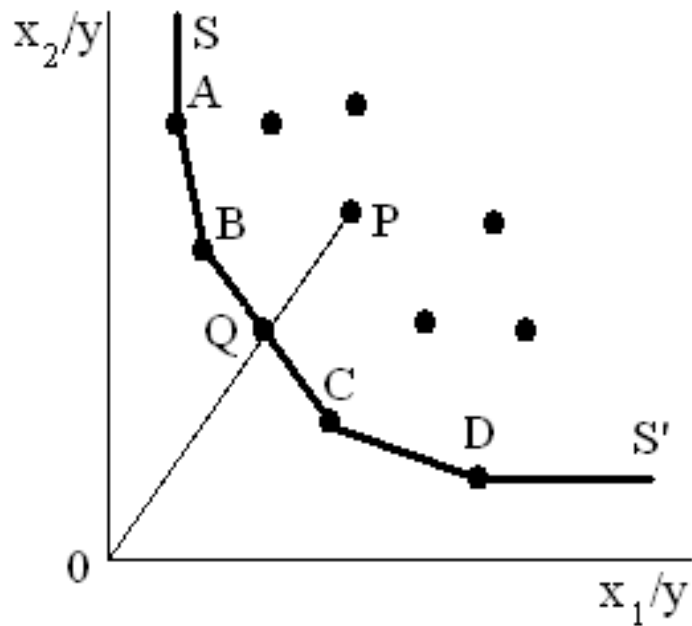
Стрелками показано направление проецирования объектов на границу эффективности (ориентация на вход или на выход)



Метод Data Envelopment Analysis

- Метод является способом оценки производственной функции
- Граница эффективности является базовым понятием метода
- Она строится в многомерном пространстве входных и выходных показателей, описывающих оцениваемые объекты
- Входные показатели – ресурсы, выходные показатели – продукция
- Степень эффективности конкретного объекта определяется расстоянием между точкой, соответствующей ему, и границей эффективности

Два входа и один выход (ориентация на вход)



- Эффективность объекта P :
$$\text{Eff} = OQ / OP$$
- A, B, C и D – эффективные объекты
- SS' – граница эффективности

Модель метода DEA (ориентация на вход)

$$\min_{\theta, \lambda} (\theta),$$

$$-\mathbf{y}_j + \mathbf{Y}\boldsymbol{\lambda} \geq \mathbf{0},$$

$$\theta \mathbf{x}_j - \mathbf{X}\boldsymbol{\lambda} \geq \mathbf{0},$$

$$\boldsymbol{\lambda} \geq \mathbf{0}.$$

дополнительное ограничение

$$\sum_{j=1}^n \lambda_j = 1$$

- n – число объектов
 m – число входных показателей
 s – число выходных показателей
- X – матрица входных показателей для всех n объектов (размерность $m \times n$)
- Y – матрица выходных показателей для всех n объектов (размерность $s \times n$)
- x_j и y_j – вектор-столбцы входных и выходных показателей для j -го – оцениваемого – объекта
- λ – вектор констант (размерность $n \times 1$)

скаляр $\theta \leq 1$ – мера (показатель) эффективности j -го объекта

Модель метода DEA (ориентация на выход)

$$\max_{\varphi, \lambda} (\varphi),$$

$$- \varphi \mathbf{y}_j + \mathbf{Y}\lambda \geq \mathbf{0},$$

$$\mathbf{x}_j - \mathbf{X}\lambda \geq \mathbf{0},$$

$$\lambda \geq \mathbf{0}.$$

дополнительное ограничение

$$\sum_{j=1}^n \lambda_j = 1$$

- n – число объектов
 m – число входных показателей
 s – число выходных показателей
- X – матрица входных показателей для всех n объектов (размерность $m \times n$)
- Y – матрица выходных показателей для всех n объектов (размерность $s \times n$)
- x_j и y_j – вектор-столбцы входных и выходных показателей для j -го – оцениваемого – объекта
- λ – вектор констант (размерность $n \times 1$)

скаляр $\varphi \geq 1$ – мера (показатель) эффективности j -го объекта



Правила применения метода DEA (1)

- Задача решается M раз (т. е. для каждого объекта):
 - если $\theta = 1$ ($\varphi = 1$), то объект эффективен;
 - если $\theta < 1$ ($\varphi > 1$), то объект неэффективен
- Неэффективные объекты можно спроецировать на границу эффективности, получив линейную комбинацию $(\mathbf{X}\lambda, \mathbf{Y}\lambda)$ – гипотетический эталонный объект



Правила применения метода DEA (2)

- Для объектов с $\theta < 1$ могут быть установлены **цели**:
пропорциональное сокращение их входных показателей в θ раз при сохранении выходных показателей на прежнем уровне
- Для объектов с $\varphi > 1$ могут быть установлены **цели**:
пропорциональное увеличение их выходных показателей в φ раз при сохранении входных показателей на прежнем уровне

Привлекательные свойства метода DEA (1)



- позволяет вычислить один агрегированный – скалярный – показатель для каждого объекта
- может одновременно обрабатывать много входов и много выходов, каждый из которых при этом может измеряться в различных единицах измерения
- позволяет учитывать внешние по отношению к рассматриваемой системе переменные – факторы окружающей среды
- не требует априорного указания весовых коэффициентов для переменных, соответствующих входным и выходным показателям при решении задачи оптимизации

Привлекательные свойства метода DEA (2)

- не налагает никаких ограничений на функциональную форму зависимости между входами и выходами
- позволяет при необходимости учесть предпочтения менеджеров, касающиеся важности тех или иных входных или выходных переменных
- производит конкретные оценки желательных изменений во входах/выходах, которые позволили бы вывести неэффективные объекты на границу эффективности
- формирует Парето-оптимальное множество точек, соответствующих эффективным объектам
- концентрируется на выявлении примеров так называемой *лучшей практики* (best practice), а не на каких-либо усредненных тенденциях, как, например, регрессионный анализ



Сферы применения метода

- государственное управление
- промышленность и сельское хозяйство
- военная сфера
- образование и здравоохранение
- транспорт
- финансовая сфера и торговля
- энергетика и энергоснабжение
- спорт
- А сфера информационных технологий?



Пример – администрирование БД

Объекты исследования – проекты по администрированию существующих баз данных. Каждый проект описывается следующими показателями:

- число человеко-часов труда специалиста высокой квалификации
- число человеко-часов труда специалиста средней квалификации
- число таблиц в базе данных
- число внешних ключей (оно отражает сложность связей в БД)
- объем базы данных (в гигабайтах)

Определение набора показателей тоже может являться отдельной задачей

Исходные данные

Номер проекта	Число человеко-часов высокой квалификации	Число человеко-часов средней квалификации	Число таблиц	Число внешних ключей	Объем базы данных (Гб)
1	5	6	5	3	2
2	5	12	6	3	6
3	6	3	12	8	1
4	9	3	7	4	4
5	2	3	4	2	1
6	5	3	5	4	3
7	6	9	7	4	5
8	2	1	6	4	1
9	5	8	5	3	1
10	8	6	10	4	2
11	3	2	6	2	5
12	5	4	6	3	4
13	4	8	15	11	2
14	7	4	6	2	4
15	4	5	4	2	3



Выбор модели

- Эффект масштаба – переменный, т. к. исходим из предположения о том, что зависимость между числом специалистов и объемом и качеством работы будет нелинейной
- Ориентация модели – на вход, т. к. повышение эффективности работы возможно за счет сокращения времени, затрачиваемого на администрирование
- В качестве *ВХОДНЫХ* показателей выбраны два показателя временных затрат
- В качестве *ВЫХОДНЫХ* показателей выбраны число таблиц, число внешних ключей и объем базы данных (Гб)

Результаты вычислений

Номер проекта	Эффективность CRS	Эффективность VRS	Эталонные проекты для данного проекта
1	0.367	0.450	11(0.250), 8 (0.750)
2	0.753	1.000	
3	0.667	1.000	
4	0.667	0.944	3 (0.167), 13 (0.111), 11 (0.722)
5	0.622	1.000	8
6	0.575	0.682	13 (0.016), 3 (0.221), 11 (0.496), 8 (0.267)
7	0.592	0.889	2 (0.500), 13 (0.167), 11 (0.333)
8	1.000	1.000	
9	0.295	0.400	8 (1.000)
10	0.413	0.530	13 (0.159), 11 (0.210), 3 (0.428), 8 (0.203)
11	1.000	1.000	
12	0.570	0.573	13 (0.067), 8 (0.200), 11 (0.733)
13	1.000	1.000	
14	0.438	0.438	11 (0.750), 8 (0.250)
15	0.505	0.625	8 (0.500), 11 (0.500)

Рекомендации для неэффективных объектов (проектов)

Проект номер 4. Уровень его эффективности равен 0.944

Показатель	Исходные значения	Рекомендуемые значения
Число высококвалифицированных человеко-часов	9.000	3.611
Число среднеквалифицированных человеко-часов	3.000	2.833
Число таблиц	7.000	8.000
Число внешних ключей	4.000	4.000
Объем базы данных (Гб)	4.000	4.000

Модель с ориентацией на вход, т. е. цель – получить рекомендации по *снижению* значений *входных* показателей. Но в ряде случаев могут выдаваться и рекомендации по *увеличению* значений *выходных* показателей. Это имеет место для показателя «Число таблиц»



Что делать с полученными результатами ?

- Гипотетический проект, который находится на границе эффективности, и будет являться целью для неэффективного проекта
- Исходят из того, что если какие-то объекты (в нашем случае – проекты) могут функционировать с высокой эффективностью, значит, и другие также *должны быть* в состоянии это делать
- Если они этого не делают, тогда нужно разбираться, в причинах: низкая квалификация специалистов, включенных в этот проект, низкая трудовая дисциплина, особенности заказчика конкретного проекта и т. д.



Пример – проектирование БД

Объекты исследования – проекты по разработке новых баз данных

Входные показатели:

- число человеко-часов труда специалиста высокой квалификации
- число человеко-часов труда специалиста средней квалификации

Выходные показатели:


- число таблиц в базе данных
- число внешних ключей (оно отражает сложность связей в БД)
- число атрибутов в таблицах

Вопрос тот же: насколько эффективно поработали наши проектировщики БД в каждом из проектов?



Обсуждение примера

- Повышение эффективности может выражаться в том, что участники проектов, не увеличивая свою численность, спроектируют более сложные базы данных (или более крупные фрагменты какой-то базы данных): содержащие больше таблиц и внешних ключей, рассчитанных на большие объемы хранимых данных и т. п.
- Следует выбирать модель с ориентацией на выход. Тогда мы получим для неэффективных проектов рекомендации по увеличению значений выходных показателей, т. е. числа таблиц и др.



Пример – оценка продуктивности работы специалистов

Можно в качестве членов выборки использовать кандидатов на вакантные должности и специалистов, уже работающих в компании

Входные показатели:

- опыт работы (в годах)

Выходные показатели:

- общий объем исходного кода, написанного программистом (число строк)
- число успешно завершенных проектов
- число языков программирования, СУБД или каких-то технологий, которыми владеет программист (как учесть уровень знаний?)

Исходные данные

Номер специалиста	Опыт работы (в годах)	Общий объем исходного кода (число строк)	Число успешно завершенных проектов	Число языков программирования, СУБД и технологий
1	18	185000	19	6
2	5	12000	3	3
3	10	57000	12	7
4	5	23000	5	3
5	8	45000	7	3
6	12	128000	16	8
7	7	37000	3	10
8	2	5000	1	3
9	32	250000	23	8
10	16	63000	15	2
11	10	71000	11	6
12	14	89000	18	5

Желтым цветом закрашены строки, соответствующие кандидатам на должность

Результаты вычислений

Номер специалиста	Эффективность CRS	Эффективность VRS	Эталонные специалисты для данного специалиста
1	0.964	1.000	
2	0.600	0.606	7 (0.100), 6 (0.250), 8 (0.650)
3	0.950	0.954	7 (0.076), 6 (0.762), 8 (0.162)
4	0.800	0.909	6 (0.300), 8 (0.700)
5	0.656	0.700	6 (0.600), 8 (0.400)
6	1.000	1.000	
7	1.000	1.000	
8	1.000	1.000	
9	0.732	1.000	
10	0.703	0.808	9 (0.111), 12 (0.889)
11	0.850	0.851	7 (0.012), 6 (0.794), 8 (0.194)
12	0.964	1.000	

Желтым цветом закрашены строки, соответствующие кандидатам на должность



Что делать с кандидатом?

Специалист номер 2. Уровень его эффективности равен 0.606

Показатель	Исходные значения	Рекомендуемые значения
Опыт работы (в годах)	5	5
Общий объем исходного кода (число строк)	12000	38950
Число успешно завершенных проектов	3	4.95
Число языков программирования, СУБД и технологий	3	4.95



Пример – оценка квалификации специалистов

В качестве членов выборки используем кандидатов на вакантные должности и специалистов, уже работающих в компании

Входные показатели:

- обобщенный входной показатель (равный 1 для всех специалистов)

Выходные показатели (оцениваются в баллах):

- уровень знаний языка SQL
- уровень знаний внутреннего устройства СУБД PostgreSQL
- уровень знаний операционной системы Linux

Исходные данные

Номер специалиста	Условный показатель затрат ресурсов	Уровень знаний языка SQL	Уровень знаний внутреннего устройства СУБД PostgreSQL	Уровень знаний операционной системы Linux
1	1	7	6	6
2	1	5	8	3
3	1	6	3	7
4	1	3	4	9
5	1	8	5	3
6	1	9	4	8
7	1	3	5	8
8	1	6	4	3
9	1	8	7	8
10	1	7	7	2
11	1	4	6	6
12	1	9	3	5

Желтым цветом закрашены строки, соответствующие кандидатам на должность

Результаты вычислений

Номер специалиста	Эффективность CRS	Эталонные специалисты для данного специалиста
1	0.871	9 (0.963), 6 (0.037)
2	1.000	
3	0.842	6 (0.687), 4 (0.312)
4	1.000	
5	0.935	6 (0.552), 9 (0.448)
6	1.000	
7	0.935	9 (0.448), 4 (0.552)
8	0.710	6 (0.455), 9 (0.545)
9	1.000	
10	0.966	2 (0.250), 9 (0.750)
11	0.837	9 (0.833), 2 (0.167)
12	1.000	6 (1.000)

Желтым цветом закрашены строки, соответствующие кандидатам на должность



Что делать с кандидатом?

Специалист номер 1. Уровень его эффективности равен 0.871

Показатель	Исходные значения	Рекомендуемые значения
Условный показатель затрат ресурсов	1	1
Уровень знаний языка SQL	7.000	8.037
Уровень знаний внутреннего устройства СУБД PostgreSQL	6.000	6.889
Уровень знаний операционной системы Linux	6.000	8.000



Рекомендации по применению метода

- Число объектов $N \geq \max \{ K \times M, 3(K + M) \}$

где K и M – числа входных и выходных переменных

- Число переменных – как правило, не более 5–6
- Для сокращения числа переменных исключать те, которые являются функционально зависящими от других переменных
- Для увеличения числа объектов в исследуемой группе можно включать в нее объекты с показателями за различные временные периоды



Программное обеспечение

- PIM-DEA Soft (Performance Improvement Management Software) (<http://deazone.com/en/software>)

Это коммерческое ПО

- DEAOS (DEA Online Software) (<https://www.deaos.com>)

Это web-приложение

- DEAP (<http://www.uq.edu.au/economics/cepa/deap.php>)

Одна из самых популярных и известных программ. Автор – австралийский профессор Т. Coelli. Эта программа является свободным ПО. Консольное приложение

- И другое ПО...



Метод DEA в России

- Первые в России – профессор В. Е. Кривоножко и его аспиранты и коллеги из Института системного анализа РАН. Их первые статьи по этому методу вышли еще в конце 90-х годов прошлого столетия

Анализ эффективности функционирования сложных систем [Текст] / В. Е. Кривоножко, А. И. Пропой, Р. В. Сеньков, И. В. Родченков, П. М. Анохин // Автоматизация проектирования. – 1999. – № 1. – С. 2–7.

- Города России, в которых «знают» о методе DEA
 - Москва
 - Санкт-Петербург (СПбГУ, Ю. В. Федотов)
 - Барнаул
 - Иваново
 - Красноярск
 - Нижний Новгород
 - Самара



Публикации в России

- Защищено более 10 диссертаций (физико-математические, технические и экономические науки)
- Статьи в журналах (в т. ч. «Экономика и математические методы»)
- Доклады на конференциях
- Учебник

Кривоножко, В. Е. Анализ деятельности сложных социально-экономических систем [Текст] / В. Е. Кривоножко, А. В. Лычев. – М. : Издательский отдел факультета ВМ и К МГУ ; МАКС Пресс, 2010. – 208 с.

Русскоязычный эквивалент названия метода

- В. Е. Кривоножко и его коллеги используют такой – «Анализ Среды Функционирования» (АСФ)

В оригинальном названии метода есть слово envelopment (обертывание). Граница эффективности как бы огибает, или обертывает, точки, соответствующие исследуемым объектам в многомерном пространстве

- «метод обволакивающей поверхности»
- «метод оболочки данных»
- «анализ свертки данных»
- «непараметрический метод анализа оболочки данных (АОД)»
- «анализ „упаковки“ (охвата) данных»



Веб-ресурсы

Самый авторитетный ресурс

<http://www.deazone.com>

Его поддерживает профессор Ali Emrouznejad

В российском сегменте Интернета аналогичного web-ресурса найти не удалось



Наш веб-ресурс

<http://www.morgunov.org/efficiency.html>

- краткое введение в метод DEA
- практический пример проведения небольшого исследования
- кандидатские диссертации авторов настоящего доклада
- доклады на конференциях и статьи, в которых рассматривается, развивается или используется метод DEA
- авторская компьютерная программа. Эта программа пока что реализует только две модели метода DEA, которые называются моделями CCR и ВСС (в их названиях используются первые буквы фамилий их авторов)

<http://www.morgunov.org/cgi-bin/dea/dea.pl>



Спасибо за внимание
