УДК 004.9:004.75

Ю.М. ЛИСЕЦКИЙ*, В.Е. СНИТЮК**

ФОРМИРОВАНИЕ ИНТЕГРАЛЬНОГО КРИТЕРИЯ ЭФФЕКТИВНОСТИ В ЗАДАЧАХ ВЫБОРА ОПТИМАЛЬНОГО ПРОЕКТНОГО ВАРИАНТА

*Институт проблем математических машин и систем НАН Украины, Киев, Украина
**Киевский национальный университет имени Тараса Шевченко, Киев, Украина

Анотація. *Розглянуто задачу вибору оптимальної альтернативи проектного варіанта при побудові корпоративної інформаційної системи. Розроблено метод формування інтегрального критерію ефективності для оцінювання альтернатив з використанням ансамблю моделей, виконано експериментальну верифікацію.*
Ключові слова: *корпоративна інформаційна система, альтернатива, критерій ефективності.*

Аннотация. *Рассмотрена задача выбора оптимальной альтернативы проектного варианта при построении корпоративной информационной системы. Разработан метод формирования интегрального критерия эффективности для оценки альтернатив с использованием ансамбля моделей, выполнена экспериментальная верификация.*
Ключевые слова: *корпоративная информационная система, альтернатива, критерий эффективности.*

Abstract. *The problem of choosing the optimal alternative for design options in building corporate information systems was reviewed. The method for forming an integral criterion for assessing the effectiveness of alternatives using assembly of models was developed, the experimental verification was performed.*
Keywords: *enterprise information system, alternative, criterion of efficiency.*

1. Введение

Известно, что информационной системой называют систему обработки информации и соответствующие организационные ресурсы (человеческие, технические, финансовые и другие), которые обеспечивают и распространяют информацию [1]. Ее неотъемлемыми компонентами являются данные, техническое, программное, информационное, лингвистическое и организационное обеспечение. Одним из видов информационных систем являются корпоративные информационные системы (КИС), которые предназначены для обслуживания информационных процессов предприятий с целью достижения их полной согласованности, безызбыточности и прозрачности, а также для поддержки принятия решений при их управлении.

Как любая сложная система, КИС состоит из множества подсистем, объединенных совокупностью разнородных связей. Приступая к созданию новой КИС, лицо, принимающее решение, неизбежно сталкивается с необходимостью осуществления выбора из многообразия объектов, элементов и взаимосвязей между ними. Очевидно, что решение задачи выбора оптимальной альтернативы зависит от спектра задач P, которые будет решать КИС, а ее эффективность будет определяться существующей структурой S и стратегией управления C, заключающейся в эффективном распределении ресурсов.

Многовариантность выбора может быть сокращена путем применения принципа двойственности погружений видов обеспечений [2] (рис. 1) в композиции с методом последовательного анализа вариантов [3]. Несмотря на значительное уменьшение количества альтернатив, мощность их множества остается значительной, что определяется разнообразием существующих технических, программных и других решений. Рассмотрим один из методов выбора оптимальной альтернативы (подсистемы, объекта, элемента и т.п.) в условиях неопределенности.

1. Методическое.
2. Организационное.
3. Информационное.
4. Программное.
5. Техническое.

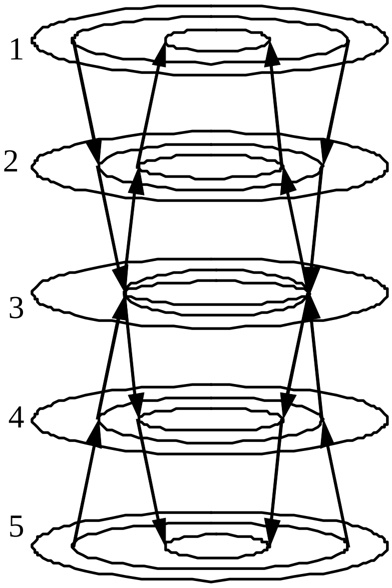

Рис. 1. Двойственность информационного погружения видов обеспечения

2. Постановка задачи

Предположим, что существует множество альтернатив $S = \{s_1, s_2, ..., s_n\}$; множество функций, которые они выполняют, $F = \{f_1, f_2, ..., f_k\}$; множество показателей эффективности выполнения каждой функции $Z = \{z_1, z_2, ..., z_k\}$. Необходимо решить задачу поиска оптимальной альтернативы

$$\underset{s \in S}{Argmax}\, E(s), \qquad (1)$$

где $E(s)$ – интегральный критерий эффективности. Очевидно, что решение задачи (1) должно быть получено с учетом ограничений, налагаемых на характеристики альтернатив. Процесс решения неизбежно имеет две составляющие: структурную и параметрическую идентификацию критерия эффективности, а также его оптимизацию.

3. Определение аддитивных критериев эффективности альтернатив

Поиск оптимальной альтернативы как процесс зависит от начальных данных. Рассмотрим его простейшие варианты. Предположим, что значения показателей эффективности для каждой альтернативы известны и сведены в матрицу $Z^* = (z_{ij})_{\mathrm{i,j}=1}^{n\ k}$, где z_{ij} – значение j-го показателя эффективности i-ой альтернативы.

Если предположить, что все значения показателей эффективности являются нормированными равнозначимыми величинами, то интегральный критерий эффективности может быть рассчитан по формуле

$$E(s_i) = \sum_{j=1}^{k} z_{ij},\ i = \overline{1,n}. \qquad (2)$$

Решение задачи (1) свелось бы к определению

$$max\{E(s_1), E(s_2), ..., E(s_n)\}. \qquad (3)$$

В практических задачах показатели эффективности являются разнозначимыми и выражение (2) приобретает такой вид:

$$E(s_1) = \sum_{j=1}^{k} w_j z_{ij},\ i = \overline{1,n}, \qquad (4)$$

где w_{j} – весовые коэффициенты, значения которых либо заданы, либо подлежат определению. Во втором случае применяют метод иерархий Саати [4], в соответствии с которым формируют матрицу попарных сравнений $A=\left(a_{ij}\right)_{i,j=1}^{n}$ и решают уравнение $Aw=\lambda w$, где λ – собственное число, w – собственный вектор. Компоненты собственного вектора w_{j}, соответствующего максимальному собственному числу, и будут искомыми весовыми коэффициентами в формуле (4), далее решают задачу (3).

Существует еще один метод определения оптимальной альтернативы без использования формул (2) и (4). Аналогично описанному выше способу находят весовые коэффициенты $w_{\mathrm{j}},\ j=\overline{1,k}$ показателей эффективности. Далее осуществляют синтез локальных приоритетов. Для этого для каждого показателя эффективности осуществляется построение матрицы попарных сравнений альтернатив. Обозначим также матрицы $A^{l}=\left(a_{ij}^{l}\right)_{i,j=1}^{n}$, $l=\overline{1,k}$. Для каждой матрицы вычисляют векторы $w^{l}=\left(w_{1}^{l},w_{2}^{l},...,w_{n}^{l}\right)$, компоненты которых будут указывать на l-ый показатель эффективности i-ой альтернативы, $i=\overline{1,n}$. Используя выражение

$$\eta_{i}=\sum_{j=1}^{k}w_{j}\cdot w_{i}^{j},\quad i=\overline{1,n} \tag{5}$$

и определяя $max\left\{\eta_{1},\eta_{2},...,\eta_{n}\right\}$, находим соответствующую оптимальную альтернативу.

Рассмотренные методы определения оптимальных альтернатив являются несложными в реализации и имеют теоретическое обоснование. Их недостаток – субъективное формирование матриц попарных сравнений, которые, к тому же, могут быть несогласованными вследствие нарушения транзитивности сравнений. В таком случае полученные результаты будут смещенными.

4. Формирование интегрального критерия эффективности системы на основе альтернативных моделей

Предположим, что исходные данные содержатся в таблице типа «объект-свойство» (табл. 1). Априори предполагаем, что критерий эффективности является неизвестной функцией от показателей эффективности, то есть

$$E=E\left(f_{1},f_{2},...,f_{k}\right). \tag{6}$$

Таблица 1. Показатели эффективности альтернатив

	f_1	f_2	–	f_k	E
s_1	z_{11}	z_{12}	–	z_{1k}	E_1
s_2	z_{21}	z_{22}	–	z_{2k}	E_2
–	–	–	–	–	–
s_n	z_{n1}	z_{n2}	–	z_{nk}	E_n

Табличные значения критерия эффективности получены в результате экспертных заключений, возможно, для аналогичных альтернатив или являются ретроспективными данными.

Какие особенности и шаги приводят к идентификации модели (6)? Во-первых, необходимо определить класс моделей, которому принадлежит искомая модель. Наиболее часто используются:

1. Модель множественной линейной регрессии:

$$E = a_0 + \sum_{i=1}^{k} a_i\, f_i \,, \tag{7}$$

где a_i – неизвестные коэффициенты. Их поиск и составляет сущность метода наименьших квадратов. Заметим, что его использование сопряжено с необходимостью тестирования и устранения мультиколлинеарности, гетероскедастичности и автокорреляции [5].

2. Модель множественной нелинейной регрессии [6]:

$$E = a \cdot \prod_{\mathrm{i}=1}^{\mathrm{k}} G_i\left(s_i\right), \tag{8}$$

где a – некоторая константа, а G_i является функцией одной переменной (показателя эффективности), которая выбирается из некоторой совокупности. Ограниченность выбора указывает на недостаток метода.

3. Модель в виде полинома Колмогорова-Габора:

$$E = a_0 + \sum_{i=1}^{k} a_i\, f_i + \sum_{i<j} a_{ij}\, f_i f_j + \sum_{i<j<l} a_{ijl}\, f_i f_j f_l + \ldots \cdot \tag{9}$$

Идентифицируют значения параметров в (9), используя метод группового учета аргументов [7]. В пользу применения такого метода свидетельствует известная теорема Вейерштрасса о сколь угодно точном приближении любой непрерывной функции полиномом. Недостатком метода группового учета аргументов является сложность аналитического построения модели (9).

4. Нейронные сети являются универсальными аппроксиматорами [8]. Для получения зависимости (6) могут быть применимы прямосвязные сети с алгоритмом обратного распространения ошибки или подобными алгоритмами обучения, сети встречного распространения, сети с радиально-базисными функциями активации и многие другие. Каждая из нейросетей является самостоятельной моделью, учет результатов их использования и знание особенностей функционирования позволяют эффективно решать задачи идентификации. К недостаткам отнесем слабую интерпретируемость нейросетей как моделей и, в отдельных случаях, зацикливание в локальных оптимумах.

5. Еще один тип моделей, которые могут быть хорошими аппроксиматорами. Это нечеткие продукционные правила [9]:

$$\text{Если } z_1 \in Z_1 \text{ и } z_2 \in Z_2, \ldots, \text{и } z_\mathrm{k} \in Z_\mathrm{k}, \text{ то } e \in E\,, \tag{10}$$

где z_i – значения показателей эффективности, Z_i – нечеткие множества с соответствующими функциями принадлежности (с неизвестными параметрами). Определение значений параметров осуществляется на основе обучения нечетких нейросетей. Недостаток, а возможно, и преимущество моделей – объективизация субъективных заключений.

Мы считаем, что для решения практических задач перечисленных моделей более чем достаточно. Все они могут быть использованы для получения (6) и для определения оптимальной альтернативы. Для этого необходимо выяснить, являются ли эти альтернативы однотипными, подобными (с экспертной точки зрения). Если это не так, то альтернативы классифицируют, используя известные методы, и табл. 1 разбивается на отдельные таблицы, соответствующие классам альтернатив. Во-вторых, таблицу необходимо разделить на две части: обучающую (А) и контрольную (В) выборки. Сделать это можно одним из способов, предложенных в [7]. Вначале строки табл. 1 упорядочиваются по убыванию их выборочных дисперсий. Далее в выборку А определяем верхние 70% строк, в выборку В – 70% нижних строк или в выборку А определяем все строки с нечетными номерами и 50% строк с четными номерами, оставшиеся строки составят выборку В. Возможны и другие процентные соотношения и способы формирования выборок. Очевидно также и то, что они будут влиять на точность результата.

На следующем шаге, используя строки обучающей выборки, осуществляем идентификацию зависимости (6) для каждой модели M_i из ансамбля моделей M, $i=\overline{1,p}$. Таким образом,

$$M_i = E_i\left(f_1, f_2, ..., f_k\right), \quad i=\overline{1,p}. \tag{11}$$

Обозначим $C=\left(C_{ij}\right)_{i,j=1}^{q\ k}$ – матрицу, соответствующую контрольной выборке и полученную в табл.1. Подставляя в модели (11) данные матрицы С, получим табл. 2.

Таблица 2. Результаты моделирования

	f_1	f_2	$\cdots$	f_k	μ_1	μ_2	$\cdots$	μ_p
s_1	C_{11}	C_{12}	$\cdots$	C_{1k}	m_{11}	m_{12}	$\cdots$	m_{1p}
s_2	C_{21}	C_{22}	$\cdots$	C_{2k}	m_{21}	m_{22}	$\cdots$	m_{2p}
$\cdots$	$\cdots$	$\cdots$	$\cdots$	$\cdots$	$\cdots$	$\cdots$	$\cdots$	$\cdots$
s_q	C_{q1}	C_{q2}	$\cdots$	C_{qk}	m_{q1}	m_{q2}	$\cdots$	m_{qp}

Значения m_{ij}, $i=\overline{1,q}$, $i=\overline{1,p}$ являются значениями критерия эффективности для i-ой альтернативы, полученными с использованием j-ой модели.

Далее, для каждой из моделей найдем показатель отклонения табличных значений критерия эффективности от значений, полученных с использованием моделей,

$$\delta_l = \frac{1}{q}\sum_{i=1}^{q}\left(\frac{m_{il} - E_i}{E_i}\right)^2, \; l=\overline{1,p}. \tag{12}$$

Значения δ_l свидетельствуют о точности аппроксимации искомой зависимости. У нас появляются основания для определения важности каждой из моделей ансамбля.

Выполним инвертирование и нормирование показателей отклонения и будем считать, что $\delta_l \in [0;1]$, $\sum_{l=1}^{p}\delta_l = 1$. Далее, для любой новой альтернативы S^* достаточно определить значения показателей эффективности z_i^*, $i=\overline{1,k}$. Подставив их в модели (11), получим значения $M_j^*, j=\overline{1,p}$. Для определения значения интегрального критерия эффективности альтернативы S^* используем формулу

$$E^* = E\left(s^*\right) = \sum_{i=1}^{p} \delta_i M_i^* \cdot \qquad (13)$$

Рассчитав значения критерия эффективности для всех альтернатив, определяем наилучшую из них. Если необходимо выполнить сравнительный анализ альтернатив, то полученные значения критерия эффективности необходимо нормировать.

Таким образом, предложенный метод может быть алгоритмизирован и алгоритм содержит такие шаги:

Шаг 1. Предварительная подготовка данных. Разделение выборки данных на обучающую и контрольную части.

Шаг 2. Для получения пула однотипных объектов выполнить кластеризацию альтернатив.

Шаг 3. Используя математические вычисления и осуществляя обучение моделей, получить критерии эффективности как зависимости от частных показателей эффективности. Используем данные обучающей выборки.

Шаг 4. Определить погрешность аппроксимации на данных контрольной выборки, которые интерпретируем как значения, обратные важности моделей.

Шаг 5. Вычисляем значения интегрального критерия эффективности для каждой альтернативы и определяем оптимальную альтернативу.

Для проверки адекватности разработанного метода проведены вычислительные эксперименты. В первом из них данные были сгенерированы по сложной многофакторной функциональной зависимости. Во втором случае исследовались статистические данные.

Для первого эксперимента получены такие коэффициенты важности моделей: δ(линейная регрессия) =0,02; δ(нелинейная регрессия)=0,14; δ(полином)=0,34; δ(перцептрон)=0,11; δ(RBF-сеть)=0,12; δ(сеть встречного распространения)=0,01; δ(нечеткие продукционные правила)=0,26. Средняя относительная ошибка на проверочных данных составила 2,3%, что более, чем в два раза меньше ошибки, полученной по методу группового учета аргументов.

Особенностью второго эксперимента являлись зашумленные данные и неочевидность зависимости результирующей характеристики от входных данных. Тем не менее коэффициенты важности имеют такие значения: δ(линейная регрессия)=0,04; δ(нелинейная регрессия)=0,12; δ(полином)=0,28; δ(перцептрон)=0,10; δ(RBF-сеть)=0,11; δ(сеть встречного распространения)=0,05; δ(нечеткие продукционные правила)=0,30. Средняя относительная ошибка проверочных данных составила 6,8%, что указывает на преимущество интегрального критерия по сравнению с наилучшей моделью – нечеткими продукционными правилами, для которых ошибка составила 9,7 % .

5. Выводы

Предложенный метод выбора варианта проектной альтернативы может быть применен с небольшими модификациями в любых задачах, связанных с необходимостью определения оптимального объекта, подсистемы, способа решения проблемы и т.п. Традиционно такого рода задачи решаются лицом, принимающим решения, исходя из опыта, интуиции и знаний. Естественно, что принятые решения являются смещенными, зачастую неэффективными, а в современных условиях и вовсе не оптимальными. Для того, чтобы уменьшить неопределенность и минимизировать доминирующий субъективизм при принятии решений, необходимо применять технологии, позволяющие объективизировать процесс определения оптимальной альтернативы.

Рассмотренная в статье технология базируется на использовании ансамбля так называемых «советующих» моделей. Определение весовых коэффициентов, указывающих на их значимость, позволяет построить критерий эффективности, являющийся интегральным показателем качества альтернатив, значения которого вычисляются без вмешательства человека.

Расширение мощности ансамбля моделей является той задачей, решение которой позволит получить усовершенствованный спектр характеристик альтернатив и уточненные показатели их качества.

СПИСОК ЛИТЕРАТУРЫ

1. ISO/IEC 2382-1:1993. Information technology. – Vocabulary. – Part 1: Fundamental terms.
2. Снитюк В.Е. Методы уменьшения неопределенности на начальных этапах проектирования систем с переменной структурой: автореф. дисс. на соискание научн. степени канд. техн. наук: спец. 05.13.12 "Системы автоматизации проектирования" / В. Е. Снитюк. – К., 1999. − 18 с.
3. Михалевич В.С. Методы последовательной оптимизации в дискретных сетевых задачах оптимального распределения ресурсов / В.С. Михалевич, А.И. Кукса. – М.: Наука, 1983. – 208 с.
4. Saaty T.L. The Hierarchon: A Dictionary of Hierarchies / Saaty T.L. – Pittsburgh, Pennsylvania: RWS Publications, 1992. – 496 p.
5. Грубер И. Эконометрия. Введение в эконометрию / Грубер И. – К.: Астарта, 1996. – Т. 1. – 434 с.
6. Снитюк В.Є. Прогнозування. Моделі, методи, алгоритми / В.Є. Снитюк. – К.: Маклаут, 2008. – 364 с.
7. Ивахненко А.Г. Моделирование сложных систем по экспериментальным данным / А.Г. Ивахненко, Ю.П. Юрачковский. – М.: Радио и связь, 1987. – 120 с.
8. Хайкин С. Нейронные сети: полный курс / Хайкин С. – М.: Вильямс, 2006. – 1104 с.
9. Kosko B. Fuzzy Systems as Universal Approximators / B. Kosko // IEEE transactions on computers. – 1994. – Vol. 43, N 11. – P. 1329 – 1333.

Стаття надійшла до редакції 05.01.2015