

УДК 658.562:001.5:519.248

**А. И. Ковалев**

ОАО ЭК «Хмельницкоблэнерго»

ул. Храновского, 11-а, 29016 Хмельницкий, Украина

## Оценивание системы управления качеством предприятия на основе кластерного анализа

*Предложен количественный метод оценивания качества функционирования предприятия, необходимый для выявления результативности осуществляемых усовершенствований. Оценивание осуществляется в рамках аудита системы управления предприятия и рассматривается как распознавание регулярностей в данных аудита (кластеров). Рассмотрено числовое представление данных аудита на примере выбора алгоритма распознавания в их фиксированных классах и построение собственно процедуры проверки. Числовое множество наблюдений аудита представлено в виде профилей данных и дендрограммы.*

**Ключевые слова:** качество, распознавание, оценка, кластерный анализ, аудит, предприятие.

Творческой стороной контроля качества функционирования предприятия (аудита) является формирование решений, которое осуществляется обычно методом эвристического поиска. Совокупность логических приемов и методических правил исследования и отыскания истины, т.е. эвристика, является полезным методом. Хотя бы потому, что сводит к минимуму или в какой-то мере ограничивает перебор возможного множества решений, сокращает время на решение по сравнению, например, с методом слепого поиска (проб и ошибок). Эвристическое исследование основывается на интуиции, находчивости, аналогиях, опыте, изобретательности. Большую роль играют накопленные знания по решению аналогичных проблем и предшествующий практический опыт. Но эвристический метод может быть неполным, неубедительным или туманным в содержании процесса получения результата или в самом результате. В эвристическом методе не требуется ни формального<sup>1</sup> определения проблемы, ни абстракции<sup>2</sup>. В целом эврис-

© А. И. Ковалев

<sup>1</sup> Формальное представление явления — это большего или меньшего масштаба логический механизм, построенный на основе формального языка и позволяющий в некоторой степени воспроизводить изучаемое явление.

<sup>2</sup> Абстракция — отвлечение от некоторых признаков явлений, не существенных для решения задачи, или отвлечение для упрощения решения на некотором этапе.

тический метод не дает возможности ясно показать, каким образом было получено заключение.

Конечно, творческие предложения, которые могут появиться в результате эвристического решения проблемы, не должны быть оставлены в стороне. Но задача состоит в том, чтобы внести структуру в слабоструктурированный процесс аудита. Надо отметить, что стандарт [1] привнес некоторую структуру посредством блок-схем: процесса управления программой аудита, типичной деятельности при проведении аудита, процесса сбора данных до принятия заключения, а также процесса поэтапного оформления заключения:

*свидетельства аудита → наблюдения аудита →  
→ заключение по результатам аудита.*

Кроме того, структура задается критериями аудита<sup>3</sup>.

Однако, как указывает автор [2], этого недостаточно. Чтобы формализовать, внести структуру в не полностью структурированный процесс решения проблемы, необходимо в последовательности этапов принятия заключения указать точки принципиальных решений, которые должны быть описаны более детально. Кроме этого, Оптнер делает ссылки на следующие требования:

— основные альтернативы<sup>4</sup> и способы их получения должны быть демонстрируемыми;

— предположения, сделанные для каждой альтернативы, должны быть определены;

— критерий, с помощью которого выносится суждение о каждой альтернативе, должен быть полностью определен (это требование нашло отражение в стандарте [1]);

— детальное представление наблюдений, взаимоотношений между наблюдениями и процедурами, с помощью которых наблюдения были оценены, должны являться частью любого заключения;

— альтернативные решения и доводы, необходимые для объяснения причин исключения отклоненных решений, должны быть доказаны.

Требование стандарта [3] принимать решения, основанные на фактах с учетом сделанных выше замечаний, побуждает искать формальные методы верификации<sup>5</sup> качества функционирования предприятия. Ниже описан метод, который не отменяет эвристический подход к отысканию истины, но добавляет некоторую

---

<sup>3</sup> *Критериями аудита* являются документы, устанавливающие требования к деятельности предприятия, а также документы, на которые делаются нормативные ссылки. *Свидетельства аудита* — записи, изложение фактов или другая информация, которые связаны с критериями аудита и могут быть перепроверены. *Наблюдения аудита* — результаты оценивания собранных свидетельств аудита относительно критериев аудита. *Заключение по результатам аудита* — выходные данные аудита, представленные группой по аудиту после рассмотрения целей аудита и всех наблюдений аудита.

<sup>4</sup> Ниже рассматривается пример на основе трех часто используемых альтернатив: состояние соответствия; состояние несоответствия; допустимое состояние, требующее улучшения.

<sup>5</sup> В общеметодологическом смысле верификация — установление фактуальности, достоверности, правдоподобности.

структуру доказательств в обычные (неформальные) рассуждения, направляемые естественно сложившимися эвристическими принципами (на основе опыта аудиторов и др.).

Указанный метод формализации предполагает, что оценка качества функционирования предприятия может быть описана и решена как задача распознавания образов. Распознавание заключается в отнесении состояния объектов аудита к одному из возможных групп объектов — кластеров  $E_q$ ,  $q=1,2,3$ . Здесь  $E_1$  — состояние соответствия (●),  $E_2$  — состояние несоответствия (○),  $E_3$  — допустимое состояние, требующее улучшения (☺). В случае процессно-функциональной организационной структуры предприятия объекты аудита — это любые комбинации подразделений, включая отделы, службы, цеха, а также творческие группы, бригады (т.е. функции) и относящиеся к ним процессы, включая операции (рис. 1). Операции — это некоторые виды деятельности в рамках процесса, каждое из которых необходимо для осуществления желаемого выхода (результата процесса), и эти действия взаимозависимы.

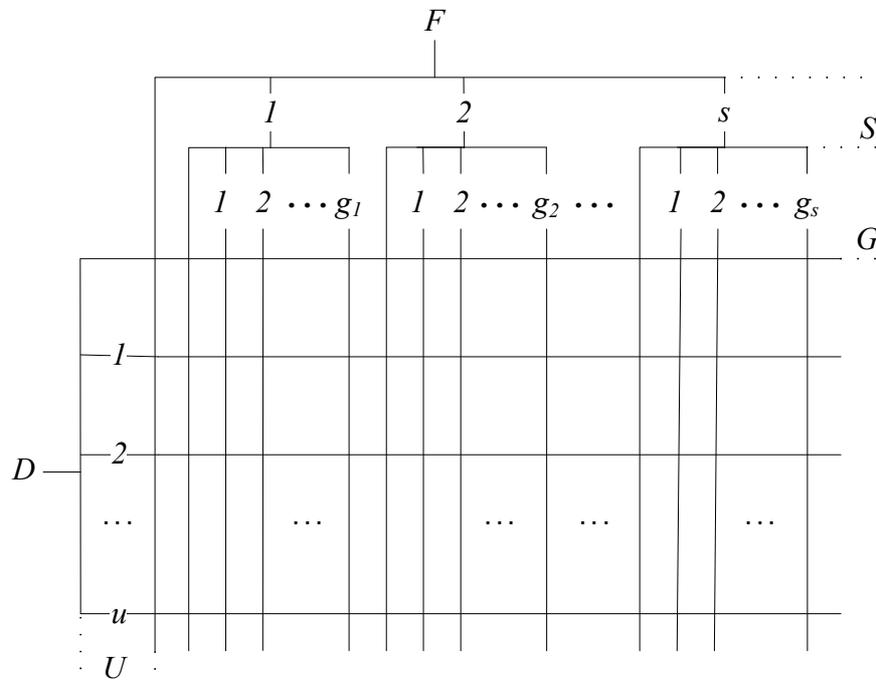


Рис. 1. Матрица объектов аудита:  $F$  — высшие руководители;  $S$  — отделы, службы, цеха;  $G$  — творческие группы, бригады;  $D$  — процессы;  $U$  — операции/этапы процессов

По поводу матрицы объектов аудита можно привести следующие пояснения. Выборочными объектами аудита являются участки на протяжении процесса, т.е. функциональные подразделения, вовлеченные в процесс. Такой аудит (аудит процессов) обеспечивает достижение целей процессов и, в конечном итоге, — целей в области качества предприятия. Потому что цели в области качества декомпозируют по целям процессов, т.е. результативные процессы приводят к результативной системе управления предприятием. Кроме проверки показателей результативности, характеризующих функционирование процессов, при аудите проверяют соблюдение требований, например стандарта [3]. Также проверяют соблюдение

прочих нормативных документов и других внешних требований (контрактов, лицензий). Исследуются также ресурсы (оборудование, материалы, люди), используемые для трансформации входов в выходы, производственная среда, применяемые методы. Кроме этого аудит процесса включает оценку последовательно осуществляемых шагов (этапов работ) и взаимосвязей с другими процессами в рамках системы. Он подразумевает анализ действий по превращению входов в выходы.

Предположим, что планируются проверки всех объектов аудита, имеющих отношение к проверяемым требованиям стандарта [3]. Для этого оформляется матрица оценивания объектов аудита (табл. 1).

Таблица 1. Матрица оценивания объектов аудита

№	Подразделение (функция)	Деятельность (процесс)
1		
2		
...	...	...
<i>N</i>		

Критерии аудита \_\_\_\_\_

№	Вопрос	Свидетельства аудита	Шкала наблюдений аудита <sup>6</sup>				
			Признаки шкалы				
			<i>p</i> <sub>1</sub>	<i>p</i> <sub>2</sub>	<i>p</i> <sub>3</sub>	...	<i>p</i> <sub><i>K</i></sub>
1	1						
	...	...				...	
	<i>T</i>						
	$\Sigma_1$		$\sigma_{11}$	$\sigma_{12}$	$\sigma_{13}$		$\sigma_{1K}$
2	1						
	...	...				...	
	<i>T</i>						
	$\Sigma_2$		$\sigma_{21}$	$\sigma_{22}$	$\sigma_{23}$		$\sigma_{2K}$
...	...	...				...	
<i>N</i>	1						
	...	...				...	
	<i>T</i>						
	$\Sigma_N$		$\sigma_{N1}$	$\sigma_{N2}$	$\sigma_{N3}$		$\sigma_{NK}$

Требования могут быть идентифицированы в шкалах наблюдений, включающих отдельные признаки (табл. 2, составляется путем структурирования требований [3]).

<sup>6</sup> Здесь должно быть приведено наименование конкретной шкалы наблюдений (например, — «Шкала управления ресурсами», или «Шкала постоянных улучшений»). На пересечении признака шкалы наблюдений и вопросов представлены собственно наблюдения аудита — результаты оценки свидетельств аудита в зависимости от критериев аудита.

Таблица 2. Шкалы наблюдений аудита. Требования ДСТУ ISO 9001

---

Шкала применения тренований:

$p_1$  — Применяемость требований.

Шкала общих требований:

$p_1$  — Разработка системы управления предприятием.

Шкала документации:

$p_1$  — Состав документации;  $p_2$  — Управление документацией;  $p_3$  — Записи по качеству.

Шкала ответственности руководства:

$p_1$  — Обязательства руководства;  $p_2$  — Фокус на потребителя;  $p_3$  — Политика в области качества;  $p_4$  — Цели в области качества;  $p_5$  — Планирование;  $p_6$  — Ответственность и полномочия;  $p_7$  — Внутренне информирование;  $p_8$  — Анализ со стороны руководства;  $p_9$  — Входные и выходные данные анализа со стороны руководства.

Шкала управления ресурсами:

$p_1$  — Обеспечение ресурсами;  $p_2$  — Компетентность персонала;  $p_3$  — Инфраструктура и производственная среда.

Шкала создания продукции:

$p_1$  — Планирование создания продукции;  $p_2$  — Процессы, связанные с потребителями.

Шкала проектирования и разработки продукции:

$p_1$  — Планирование и управление проектированием;  $p_2$  — Входные и выходные данные проектирования и разработки;  $p_3$  — Анализ проектирования и разработки;  $p_4$  — Проверка и утверждение проекта (разработки).

Шкала закупок:

$p_1$  — Процесс закупок;  $p_2$  — Информация по закупкам;  $p_3$  — Проверка закупленной продукции.

Шкала производства и предоставления услуг:

$p_1$  — Управление производством и предоставлением услуг;  $p_2$  — Идентификация и прослеживаемость;  $p_3$  — Собственность потребителей;  $p_4$  — Сохранение продукции.

Шкала управления средствами измерительной техники:

$p_1$  — Управление средствами измерительной техники.

Шкала мониторинга и измерений:

$p_1$  — Осуществление мониторинга и измерений;  $p_2$  — Удовлетворенность потребителей;  $p_3$  — Внутренние аудиты;  $p_4$  — Мониторинг и измерение процессов;  $p_5$  — Мониторинг и измерение продукции.

Шкала управления несоответствующей продукцией:

$p_1$  — Управление несоответствующей продукцией.

Шкала анализа данных:

$p_1$  — Осуществление анализа данных;  $p_2$  — Управление анализом данных.

Шкала постоянных улучшений:

$p_1$  — Осуществление постоянных улучшений;  $p_2$  — Корректирующие действия;  $p_3$  — Препьюждающие действия.

---

Необходимо адекватно оценить выполнение установленных требований и определить объекты потенциального улучшения по результатам анализа наблюдений аудита. Анализ наблюдений аудита предполагает адекватную оценку объектов аудита по оценкам множества частных вопросов. Вопросы оформляются в виде опросного листа и применяются ко всем объектам аудита. Ответ на каждый из вопросов позволяет: а) установить свидетельство аудита; б) отнести соответст-

вующий объект аудита к одному из признаков шкалы наблюдений, а именно к тому, для которого идентифицирована возможность улучшения.

Анализ на основе  $T$  стандартных вопросов представляет диагностический тест, который можно назвать многофазным обследованием объектов аудита (МООА). МООА-тест применим к однородным, в некотором смысле, объектам аудита. Однородными, т.е., к которым применим один тест, могут быть объекты аудита, стоящие на пересечении одного из  $s$  столбцов и одной из  $u$  строк или строки  $D$  (рис. 1). Например, это — отделы конструкторского бюро, бригады механообрабатывающего производства, экипажи самолетов, вовлеченные в один процесс. В рамках аудита однородной группы объектов возможно использование нескольких шкал, одной шкалы или части шкал. Главное требование — рассмотрение всех шкал в течение годового цикла проверки предприятия. В теории распознавания образов конкретные значения шкал наблюдений называют признаками.

Эмпирическим путем  $T$  вопросов группируются по признакам шкалы МООА-теста. Аудитор относит вопрос к тому или иному признаку  $p_j$ ,  $j = 1, 2, \dots, K$ , если он (вопрос) позволяет идентифицировать улучшение по этому признаку (т.е. имело место несоответствие). Тогда в матрице оценивания на пересечении соответствующего вопроса и признака ставится знак «1». Количество знаков «1» суммируется для каждого признака соответствующего объекта аудита ( $\sigma_{lj}$ ;  $l = 1, 2, \dots, N$ ;  $j = 1, 2, \dots, K$ ). МООА-тест предлагается одному человеку, или нескольким (группе), относящимся к проверяемому объекту аудита. В случае опроса группы, вычисляются средние для группы значения  $\sigma_{lj}$ .

Таким образом, объекты аудита могут быть представлены точками  $\Sigma_l = \{\sigma_{lj}\}$ ,  $l = 1, 2, \dots, N$ ;  $j = 1, 2, \dots, K$  в  $K$ -мерном числовом пространстве признаков. Возможна графическая интерпретация МООА-теста с помощью метода параллельных координат, который основан на конструировании двумерных геометрических моделей (профилей данных аудита) для объектов  $K$ -мерного евклидова пространства [4]. На плоскости с декартовыми координатами  $p\sigma$  располагаются  $K$  копий оси ординат  $\sigma$  (ось численных значений признаков) на равных расстояниях перпендикулярно оси абсцисс  $p$  (ось признаков шкалы наблюдения) и обозначаются  $p_1, p_2, \dots, p_K$  (рис. 2).

Формально профиль есть вектор значений признаков, графически изображаемый в виде ломаной линии. Каждый объект аудита имеет свой профиль — своеобразную «кардиограмму» состояния объекта. Сходство между профилями объектов аудита определяют три элемента: *форма*, т.е. спуски и подъемы ломаной линии для всех признаков; *рассеяние*, т.е. дисперсия значений признаков относительно их среднего; *уровень* (или сдвиг), т.е. среднее значение для объекта аудита по всем признакам [5]. В первую очередь профили различаются по признакам, которые имеют наивысшее значение (пики).

Для интерпретации первичных данных аудита, т.е. наблюдений, введем в рассмотрение матрицу размерностью  $N \times K$ , состоящую из  $N$  (по числу объектов аудита) строк и  $K$  (по числу признаков шкалы) столбцов:

$$\Sigma = \begin{bmatrix} \sigma_{11}, & \dots, & \sigma_{1K} \\ \sigma_{21}, & \dots, & \sigma_{2K} \\ \dots & \dots & \dots \\ \sigma_{N1}, & \dots, & \sigma_{NK} \end{bmatrix}, \quad (1)$$

где  $\sigma_{lj}$  — значение  $j$ -го признака, подсчитанного для  $l$ -го объекта аудита.

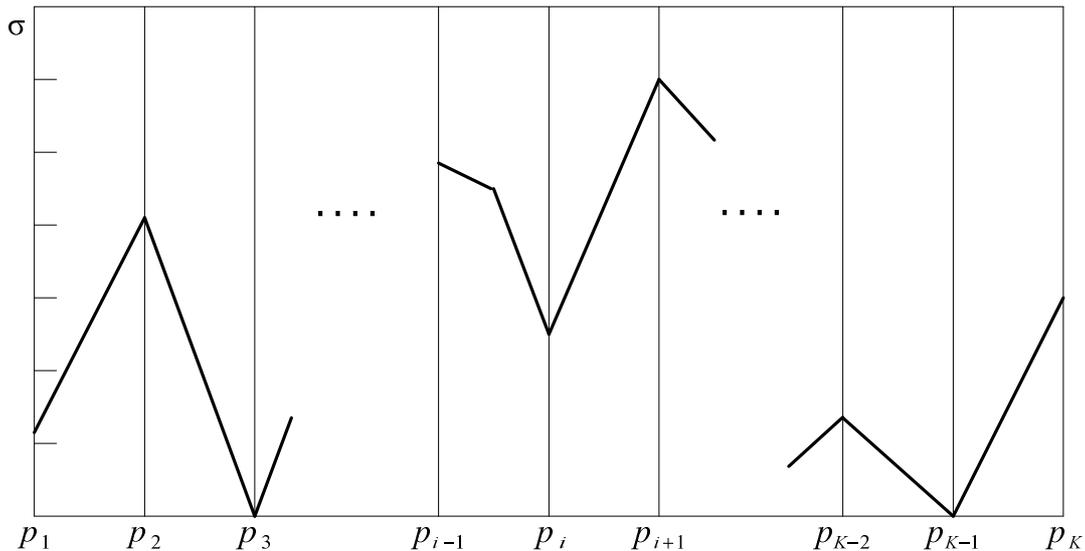


Рис. 2. Пример профиля данных МООА-теста

Таким образом, строка  $\Sigma_l = (\sigma_{l1}, \sigma_{l2}, \dots, \sigma_{lK})$  является совокупностью значений  $K$  признаков, вычисляемых у  $l$ -го объекта аудита. Столбец  $(\sigma_{1j}, \dots, \sigma_{Nj})'$  — это совокупность  $N$  значений  $j$ -го признака, рассчитанного на множестве  $N$  объектов аудита. Далее объектом будем называть саму строку матрицы данных (оценивания).

В пространстве описания каждый объект аудита представляется  $K$ -мерным вектором или точкой, координаты которой есть значения соответствующих признаков. Исходную выборку (1) требуется разбить на однородные в некотором смысле кластеры — непрерывные области некоторого пространства с относительно высокой плотностью точек, отделенные от других таких же областей областями с относительно низкой плотностью точек. Пользуясь кластер-анализом, аудитор пытается определить естественное расслоение объектов исходной выборки на более или менее ярко выраженные кластеры, лежащие друг от друга на некотором расстоянии в многомерном пространстве описания.

Наибольшую трудность представляет определение того, что понимать под группой или кластером. Сделать это можно разными способами, но почти все они основаны на понятии «близости», или «расстояния» между объектами. Следовательно, необходимо ввести некоторую метрику  $r(\Sigma_l, \Sigma_m)$ , определяющую рас-

стояние между двумя точками  $\Sigma_l$  и  $\Sigma_m$ , а затем установить расстояние, при котором два объекта аудита можно считать близкими. Очевидно, что чем меньше расстояние между объектами, тем они ближе.

Матрица (1) задает отношения «объект-признак». По этим первичным данным вычисляется матрица сходства объектов аудита с помощью меры сходства. Это квадратная симметричная матрица размером  $N \times N$ , задающая отношение «объект-объект»:

$$R = \begin{bmatrix} r_{11}, & \dots, & r_{1N} \\ r_{21}, & \dots, & r_{2N} \\ \dots & \dots & \dots \\ r_{N1}, & \dots, & r_{NN} \end{bmatrix}. \quad (2)$$

Элементы  $r_{lm}$  этой матрицы являются значениями некоторой меры сходства между объектами  $\Sigma_l$  и  $\Sigma_m$  [6]. Матрица  $R$  вычисляется по матрице  $\Sigma$  в качестве промежуточного результата при применении процедур кластерного анализа (распознавания). Проблема распознавания заключается в том, чтобы анализируемую совокупность объектов аудита  $\Sigma_1, \dots, \Sigma_N$ , представленную в виде (1), разбить (классифицировать) на небольшое число однородных в определенном смысле групп. Говоря о классификации совокупности объектов, будем подразумевать, что каждый из них задан соответствующей строкой матрицы (1), а геометрическая структура их попарных расстояний (связей) задана матрицей (2). Мы ожидаем, что алгоритм распознавания позволит разбить множество объектов аудита на группы  $E_q$ ,  $q=1,2,3$ , причем в группу  $E_2$  (несоответствия) попадут объекты с наибольшими числовыми оценками признаков, т.е. для которых выявлено большее число возможностей для улучшения. Для формализации этой проблемы, как было сказано выше, удобно интерпретировать анализируемые объекты в качестве точек в факторном пространстве признаков. Если исходные данные представлены в виде (1), то эти точки являются непосредственным геометрическим изображением многомерных наблюдений  $\Sigma_l = (\sigma_{l1}, \sigma_{l2}, \dots, \sigma_{lK})$  в  $K$ -мерном пространстве признаков. Если же исходные данные удастся представить в виде (2), то тем самым задана структура попарных расстояний между объектами. Выбор меры сходства и метода разбиения является важным этапом решения задачи распознавания состояния проверяемой группы объектов аудита. Причем выбранная совокупность признаков  $p_j$  сама в определенной мере отражает понятие сходства объектов — в смысле соответствия (несоответствия) сформулированным требованиям [7]. Решение задачи может быть не единственным, поэтому его будем рассматривать не как структуру, являющуюся объективным свойством данных, а как структуру для анализа аудиторами.

Самой распространенной мерой сходства для приложений, в которых данные описываются в терминах профилей, является коэффициент корреляции [5]

$$r_{lm} = \frac{\sum_{i=1}^K (\sigma_{li} - \bar{\sigma}_l)(\sigma_{mi} - \bar{\sigma}_m)}{\sqrt{\sum_{i=1}^K (\sigma_{li} - \bar{\sigma}_l)^2 \sum_{i=1}^K (\sigma_{mi} - \bar{\sigma}_m)^2}}; \quad l, m = 1, 2, \dots, N, \quad (3)$$

где  $\sigma_{li}, \sigma_{mi}$  — значения  $i$ -го признака для  $l$ -го и  $m$ -го объектов аудита соответственно;  $\bar{\sigma}_l, \bar{\sigma}_m$  — средние всех значений признаков  $l$ -го и  $m$ -го объектов аудита соответственно.

Ценность корреляции в том, что она оценивает форму, т.е. не зависит от различий между признаками из-за рассеяния и уровня. Ибо можно предположить, что влияние рассеяния и уровня данных будет определяться субъективизмом аудиторов, а не недостатками, присущими классификации.

Естественно предполагать, что геометрическая близость точек  $\Sigma_1, \dots, \Sigma_N$  в  $K$ -мерном пространстве признаков обозначает близость фактических состояний соответствующих объектов аудита. Тогда проблема классификации состоит в разбиении анализируемой совокупности точек — наблюдений на сравнительно небольшое число классов, таких, что точки, принадлежащие к одному классу, «близки» друг к другу, а точки из разных классов «далеки» друг от друга.

При конструировании различных кластер-процедур очень часто вводят понятие расстояния не только между отдельными объектами, а между целыми группами объектов. Понятие расстояния между группами объектов особенно важно в так называемых иерархических кластер-процедурах, где алгоритм последовательно объединяет (или разъединяет) самые близкие (далекие) друг от друга объекты, а затем и целые группы объектов. Различают агломеративные и дивизимные кластер-процедуры. Агломеративные иерархические кластер-процедуры представляют собой пошаговый алгоритм, при котором на каждом шаге происходит объединение множества объектов, подлежащих классификации на непересекающиеся кластеры, при этом каждое последующее объединение относится к кластерам, полученным на предыдущем шаге. Дивизимные кластер-процедуры имеют противоположную направленность: на каждом шаге дробят полученные ранее кластеры на более мелкие. Иерархические классификации подразделяют также на монотетические и политетические. В монотетической классификации деление производится на основании одного признака, имеющего максимальную информативность, тогда как в политетической классификации все признаки учитываются в равной степени. В нашей задаче интерес представляют иерархические агломеративные кластер-процедуры — они всегда политетические. Эти процедуры привлекательны тем, что делают полный и достаточно тонкий анализ структуры исследуемого множества объектов. Также они дают возможность наглядной интерпретации последовательного объединения наиболее схожих объектов в виде иерархического дерева (дендрограммы). Общая схема иерархической агломеративной процедуры имеет следующий вид:

- 1) все объекты считаются отдельными кластерами;
- 2) два самых близких кластера (по матрице сходства, т.е. межклассовых расстояний) объединяются в один;

- 3) пересчитывается матрица сходства;
- 4) переход к п. 2.

Такой алгоритм за  $(N - 1)$  шагов объединяет все объекты в один кластер.

Начальные действия во всех агломеративных процедурах одинаковы. Для  $N$  объектов вычисляются все  $N \cdot (N - 1) / 2$  мер сходства и пара объектов с наименьшей мерой объединяется в один кластер. Необходимо затем определить подходящую меру сходства между этим кластером и остальными  $N - 2$  объектами, а на более поздних стадиях, очевидно, будет необходимо определить меру между объектом и кластером любого объема, а также между любыми двумя кластерами. На каждом шаге классификации осуществляется то объединение (между двумя объектами, между объектом и кластером или между двумя кластерами), для которого мера различия минимальна среди всех оставшихся к данному шагу. Мера должна быть такой, чтобы объект мог рассматриваться как кластер из одного объекта. Процедура объединения определяется именно мерой различия между кластерами. Рассмотрим некоторые меры, т.е. расстояния, характеризующие взаимное расположение отдельных кластеров. Каждому расстоянию соответствует своя стратегия объединения. Приведем наиболее употребительные.

1. *Стратегия ближайшего соседа.* Расстояние между двумя кластерами определяется в этом случае как расстояние между ближайшими объектами из этих кластеров.

2. *Стратегия дальнего соседа.* В этой стратегии расстояние между двумя кластерами определяется как расстояние между двумя самыми удаленными представителями этих кластеров.

3. *Стратегия группового среднего.* Если один кластер состоит из  $n_1$ , а другой из  $n_2$  объектов, то расстояние между этими кластерами в данной стратегии определяется как среднее арифметическое из  $n_1 \cdot n_2$  попарных расстояний между объектами из этих кластеров, т.е. это расстояние определяется как среднее арифметическое всевозможных попарных расстояний между представителями рассматриваемых кластеров.

4. *Гибкая стратегия.* Это комбинаторная стратегия, свойства которой полностью зависят от выбора так называемого коэффициента экспансии, — коэффициент  $\beta$  в общей формуле (4) для вычисления расстояния между кластерами (см. ниже).

Следует отметить, что каждая конкретная стратегия описывает в исходной матрице оценивания не те кластеры, которые там реально существуют, а те, для поиска которых она предназначена. Этот факт является следствием того, что круг задач, при решении которых не происходит принудительного навязывания наблюдениям аудита заранее predetermined структуры, для конкретной стратегии достаточно узок.

Приведем анализ областей применения поименованных выше стратегий объединения (и соответствующих им функций межклассового расстояния).

Напомним, мы рассматриваем объекты аудита как точки в многомерном пространстве, оси которого соответствуют признакам. Некоторые стратегии объединения устанавливают границы между группами точек, но не изменяют относительного расположения точек в пространстве. Такие стратегии называют страте-

гиями, сохраняющими метрику пространства. В других стратегиях пространство вокруг группы растягивается по мере того, как группа растет, так что кажется, что группа отстывает от остальных точек. Такие стратегии называются *растягивающими пространство*. Они интенсифицируют классификацию, и получающиеся группы кажутся более различающимися, чем это имеет место в действительности. Есть также стратегии, в которых пространство как бы сжимается вокруг группы по мере ее роста (*сжимающие пространство стратегии*). Естественная кластеризация в этом случае «размывается», и может появиться множество «цепочек», состоящих из последовательно соединившихся друг с другом объектов.

Пусть имеются две группы  $g$  и  $h$  с  $n_g$  и  $n_h$  объектами соответственно; мера сходства между этими группами обозначается  $r_{g,h}$ . Допустим, что  $r_{g,h}$  — минимальная мера из всех оставшихся, так что  $g$  и  $h$  объединяются и образуют новую группу  $e$  с  $n_e = n_g + n_h$  объектами. Рассмотрим некоторую другую группу  $c$  с  $n_c$  объектами. Перед объединением известны значения  $r_{c,g}$ ,  $r_{c,h}$ ,  $r_{g,h}$ ,  $n_c$ ,  $n_g$ ,  $n_h$ . Положим

$$r_{c,e} = \alpha_g \cdot r_{c,g} + \alpha_h \cdot r_{c,h} + \beta \cdot r_{g,h} + \gamma \cdot |r_{c,g} - r_{c,h}|, \quad (4)$$

где параметры  $\alpha_g$ ,  $\alpha_h$ ,  $\beta$ ,  $\gamma$  определяют сущность стратегии и конечный вид группировки.

Для того чтобы графически представить весь процесс объединения, объекты должны быть размещены в соответствующем порядке вдоль оси абсцисс. Но последовательность объединений (иерархия или дендрограмма) требует, чтобы каждое объединение было связано с некоторым значением ординаты. Обычно для этого используют меру различия, соответствующую данному объединению. Дендрограмма будет очень ненаглядной, если она не будет расти при каждом объединении, что соответствует требованию, чтобы ряд мер различия, связанных с последовательными объединениями, был монотонным. Если  $\gamma \neq 0$ , то все известные комбинаторные стратегии, реализующие правило (4), монотонны; остальные стратегии монотонны, если  $\alpha_g + \alpha_h + \beta \geq 1$ .

Возвращаясь к перечисленным выше стратегиям, отметим, что стратегия ближайшего соседа реализуется при  $\alpha_g = \alpha_h = 0,5$ ;  $\beta = 0$ ;  $\gamma = -0,5$ . Эта стратегия монотонна, она сильно сжимает пространство, что делает ее неподходящей во многих стандартных задачах классификации. Стратегия дальнего соседа реализуется при  $\alpha_g = \alpha_h = 0,5$ ;  $\beta = 0$ ;  $\gamma = 0,5$ . Это монотонная, сильно растягивающая пространство стратегия, которая в значительной степени может быть заменена гибкой стратегией. Стратегия группового среднего реализуется, если параметры равны:  $\alpha_g = n_g/n_e$ ,  $\alpha_h = n_h/n_e$ ,  $\beta = \gamma = 0$ . Стратегия монотонна и сохраняет метрику пространства. Она рекомендуется, когда искусственное разделение групп не нужно.

Гибкая стратегия применима для любой меры различия и определяется четырьмя ограничениями:

$$\alpha_g + \alpha_h + \beta = 1; \quad \alpha_g = \alpha_h; \quad \beta < 1; \quad \gamma = 0.$$

Стратегия монотонна и ее свойства полностью зависят от коэффициента  $\beta$ , который иногда называют коэффициентом экспансии. Если  $\beta = 0$ , то стратегия сохраняет метрику пространства. Если  $\beta > 0$ , то стратегия сжимает пространство, в этом случае кластер-процедура оказывается наиболее приспособленной для поиска значительного числа более или менее однородных (в смысле расстояний между объектами) групп. Если  $\beta < 0$ , то стратегия растягивает пространство, в этом случае иерархическая кластер-процедура оказывается ориентированной на поиск малого числа групп объектов, имеющих явно выраженные сгущения. На практике обычно используют значение  $\beta = -0,25$ ; рекомендуемый диапазон:  $-0,25 \div 0,25$ .

Рассмотрим пример, в котором используется гибкая стратегия. Предположим, что имеется девять объектов аудита, для которых рассчитана следующая матрица сходства вида (2):

	$\Sigma_1$	$\Sigma_2$	$\Sigma_3$	$\Sigma_4$	$\Sigma_5$	$\Sigma_6$	$\Sigma_7$	$\Sigma_8$	$\Sigma_9$
$\Sigma_1$	–	0,773	0,750	0,578	0,103	0,500	0,817	0,420	0,245
$\Sigma_2$		–	0,508	0,613	0,281	0,500	0,625	0,660	0,405
$\Sigma_3$			–	0,644	0,410	0,375	0,901	0,345	0,509
$\Sigma_4$				–	0,327	0,475	0,698	0,540	0,109
$\Sigma_5$					–	0,525	0,562	0,701	0,220
$\Sigma_6$						–	0,795	0,655	0,300
$\Sigma_7$							–	0,480	0,411
$\Sigma_8$								–	0,550
$\Sigma_9$									–

В расчетах используем  $\alpha_g = \alpha_h = 0,625$ . Наибольшее значение в матрице равно  $r_{3,7} = 0,901$ . Объекты 3 и 7 объединяются и образуют группу 10 (рис. 3). Вычисляются значения  $r_{f,10}$ , где  $f$  — каждый из оставшихся объектов:  $r_{1,10}; r_{2,10}; r_{4,10}; r_{5,10}; r_{6,10}; r_{8,10}; r_{9,10}$ .

$$r_{1,10} = 0,625 \cdot (r_{1,3} + r_{1,7}) - 0,25 \cdot r_{3,7} = 0,625 \cdot (0,750 + 0,817) - 0,25 \cdot 0,901 = 0,754;$$

$$r_{2,10} = 0,625 \cdot (r_{2,3} + r_{2,7}) - 0,25 \cdot r_{3,7} = 0,625 \cdot (0,508 + 0,625) - 0,25 \cdot 0,901 = 0,483;$$

$$r_{4,10} = 0,625 \cdot (r_{4,3} + r_{4,7}) - 0,25 \cdot r_{3,7} = 0,625 \cdot (0,644 + 0,698) - 0,25 \cdot 0,901 = 0,613;$$

$$r_{5,10} = 0,625 \cdot (r_{5,3} + r_{5,7}) - 0,25 \cdot r_{3,7} = 0,625 \cdot (0,410 + 0,562) - 0,25 \cdot 0,901 = 0,383;$$

$$r_{6,10} = 0,625 \cdot (r_{6,3} + r_{6,7}) - 0,25 \cdot r_{3,7} = 0,625 \cdot (0,375 + 0,795) - 0,25 \cdot 0,901 = 0,506;$$

$$r_{8,10} = 0,625 \cdot (r_{8,3} + r_{8,7}) - 0,25 \cdot r_{3,7} = 0,625 \cdot (0,345 + 0,480) - 0,25 \cdot 0,901 = 0,290;$$

$$r_{9,10} = 0,625 \cdot (r_{9,3} + r_{9,7}) - 0,25 \cdot r_{3,7} = 0,625 \cdot (0,509 + 0,411) - 0,25 \cdot 0,901 = 0,350.$$

После того, как кластеры  $g$  и  $h$  (на данном шаге — это объекты 3 и 7) объединены в новый кластер  $e$  (на данном шаге — 10) и вычислены значения  $r_{f,10}$ , старые векторы  $g$  и  $h$  оказываются ненужными, и новый вектор может быть записан на место одного из них. Поэтому столбец 3 и строка 3 выбрасываются, вместо столбца 7 и строки 7 записываются столбец 10 и строка 10. Получается:

	$\Sigma_1$	$\Sigma_2$	$\Sigma_4$	$\Sigma_5$	$\Sigma_6$	10	$\Sigma_8$	$\Sigma_9$
$\Sigma_1$	–	0,773	0,578	0,103	0,500	0,754	0,420	0,245
$\Sigma_2$		–	0,613	0,281	0,500	0,483	0,660	0,405
$\Sigma_4$			–	0,327	0,475	0,613	0,540	0,109
$\Sigma_5$				–	0,525	0,383	0,701	0,220
$\Sigma_6$					–	0,506	0,655	0,300
10						–	0,290	0,350
$\Sigma_8$							–	0,550
$\Sigma_9$								–

Наибольшее значение в матрице сходства теперь равно  $r_{1,2} = 0,773$ . Объединяя объекты 1 и 2, получаем группу 11 и вычисляем  $r_{4,11}; r_{5,11}; r_{6,11}; r_{10,11}; r_{8,11}; r_{9,11}$ . Действуя аналогичным образом, получаем группы 15 и 16 и вычисляем значение  $r_{15,16}$ , которое равно:

$$r_{15,16} = 0,625 \cdot (r_{15,14} + r_{15,9}) - 0,25 \cdot r_{14,9} = (0,177 + 0,078) - 0,25 \cdot 0,238 = 0,099.$$

Объединение 15 и 16 дает группу 17. Поскольку осталась только одна группа, расчет прекращается. Соответствующая дендрограмма приведена на рис. 3. Значения подобия между группами (вертикальная ось — коэффициенты корреляции) используют для их объединения. Из дендрограммы следует, что объединения 10, 11, 13, 15, а также 12, 14 — естественные, но значительное уменьшение подобия, необходимое для перехода на следующий уровень, делает объединение 15 и 16, а также 14 и 9 вынужденным. Дендрограмма на рис. 3 показывает, что найденное решение состоит из трех кластеров (третий кластер состоит из одного объекта, — это  $E_2$  — состояние несоответствия). Существуют формальные критерии, позволяющие определить число полученных кластеров и методы проверки достоверности результатов. Их использование предваряется анализом применимости, достоинств (недостатков). Однако оценка качества и разумность полученного разбиения должна носить, прежде всего, содержательный характер. Формальные критерии оказывают лишь помощь в этом вопросе. Целесообразно применить несколько мер сходства и алгоритмов классификации с последующим сравнением результатов. Таким путем можно найти полезный метод, который будет в согласии с содержанием решаемой задачи, применяемыми признаками и мерой сходства. По-

лучаемые при этом частично обучающие выборки (от 3–5 % общего числа объектов аудита) позволяют значительно повысить эффективность классификации [6].

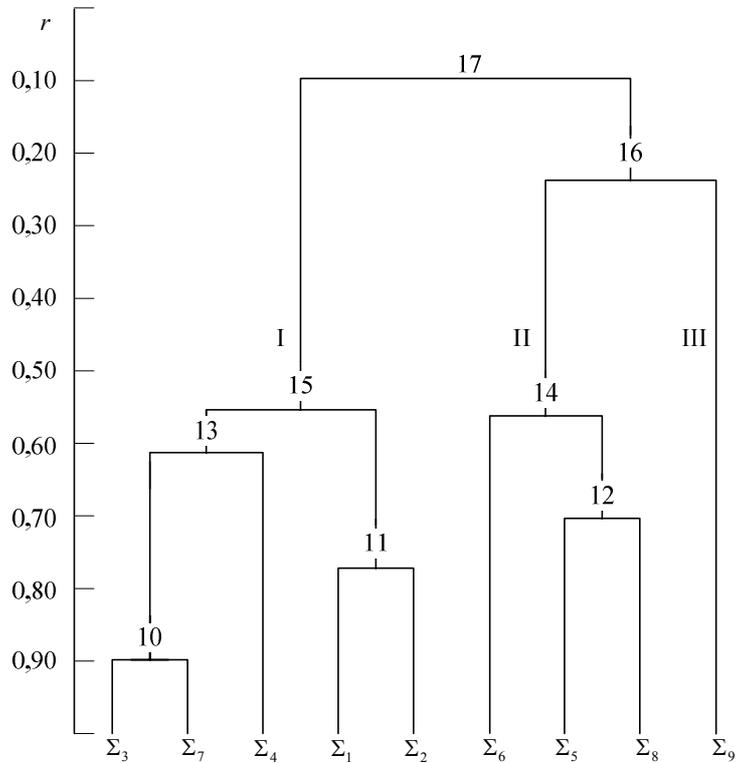


Рис. 3. Дендрограмма иерархической группировки

В табл. 3 приведена форма заключения по результатам аудита.

Таблица 3. Заключение по результатам аудита

№	Подразделение (функция)	Деятельность (процесс, процедура)
1		
2		
...	...	...
$N$		
Цели аудита _____		
Выводы аудита <sup>7</sup>		Рекомендации по улучшению
№	$E$	
1	.	
2	.	
...	...	...
$N$	.	

<sup>7</sup> На основании процедуры кластеризации:  $E_1 = \bullet$ ,  $E_2 = \circ$ ,  $E_3 = \odot$

Выводы аудита дополняются анализом наблюдений (т.е. профилей) с целью: а) согласования выводов; б) формулировки рекомендаций о возможных улучшениях. При этом рекомендации об улучшении объектов аудита будут формулироваться в первую очередь на основе рационально выбранных параметров, характеризующих кластер. Такими параметрами являются, например, центр кластера, мера вытянутости (расстояние между наиболее удаленными друг от друга, но принадлежащими одному кластеру объектами), интегральная плотность (число признаков в каждой шкале наблюдений).

Полезно визуальное представление кластера по аналогии с профилем отдельного объекта аудита как точки в  $K$ -мерном пространстве признаков (рис. 2), но в виде семейства профилей — как отрезка прямой в пространстве признаков. На рис. 4 представлено семейство профилей объектов аудита, имеющих отношение к шкале мониторинга и измерений (см. табл. 2). Эти процессы включают, в том числе, оценивание удовлетворенности потребителей (признак  $p_2$ ), осуществляемое посредством аудитов потребителей. Аудит службы качества предприятия осуществляется представителями высшего руководства.

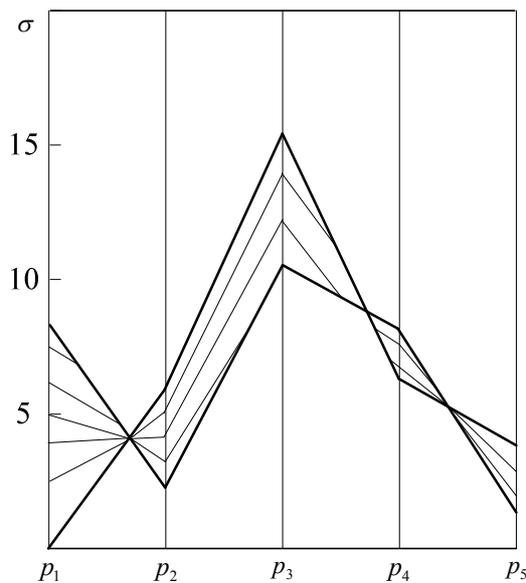


Рис. 4. Семейство профилей шкалы мониторинга и измерений для кластера  $E_3$  — допустимое состояние, требующее улучшения

Рис. 4 построен для части объектов аудита, попавших в группу  $E_3$  (допустимое состояние, требующее улучшения). Как видно, наибольшее число замечаний по улучшению касается внутренних аудитов ( $p_3$ ), причем наблюдается значительный разброс (вариабельность) значений признаков. Этот разброс определяется несвоевременным (в ряде подразделений) принятием мер по устранению несоответствий и почти повсеместное отсутствие проверок предпринятых мер (отсутствуют отчеты о результатах проверки). Существенный разброс оценок по признаку  $p_1$  (осуществление мониторинга и измерений) связан с различной степенью

применимости (от процесса к процессу) статистических методов, отсутствием, в некоторых случаях, измерений важных параметров техпроцессов и отсутствием перехода от первичных результатов измерения к некоторым обобщенным показателям результативности. Относительно большая доля замечаний по признаку  $p_4$  (мониторинг и измерение процессов) свидетельствовала, в основном, о том, что критерии измерения процессов были случайные и не указывали явно на достижение запланированных результатов.

## Заключение

Формализованное описание аудита позволяет представить большое число частных факторов с помощью небольшого числа основных результатов. Каждая формализация фиксирует лишь нечеткий срез всегда очень сложной реальности. Но при этом она дает возможность уменьшить избыточность запоминаемой информации с целью познания закономерных связей. А также уменьшить элемент неопределенности заключения по результатам аудита. Формализация является существенным вспомогательным средством отыскания решений. Представленный метод характеризует следующие основные шаги:

- 1) в соответствии с планом аудита определяются объекты для анализа и кластеризации (рис. 1);
- 2) в качестве требований устанавливаются шкалы наблюдений, по которым будут оценивать объекты в выборке (табл. 2);
- 3) оценивается выполнение требований посредством стандартного МООА-теста и вычисляется значение меры сходства между объектами (табл. 1);
- 4) применяется кластерный анализ для распознавания групп схожих объектов (рис. 3);
- 5) для объектов каждой группы анализируется семейство профилей признаков с целью формулировки рекомендаций по улучшению (рис. 4), оформляется заключение по результатам аудита (табл. 3).

Значимость несоответствий может быть представлена количеством замечаний о возможных улучшениях — суммой единиц, полученных в матрице оценивания по каждой шкале и для каждого объекта. Таким образом, аудитор освобожден от необходимости постоянно анализировать важность выявленного несоответствия. При достаточно большом числе вопросов существенные несоответствия сами обратят на себя внимание — через большее число возможностей для улучшения.

Существенной особенностью предложенного метода является возможность непрерывного совершенствования в смысле его полезности, — по мере использования новых данных (обучающие выборки). Речь идет о создании банков данных в форме профилей, которые можно запрашивать с различными целевыми установками. Стандартные условия проверок (МООА-тест) обеспечивают хорошую воспроизводимость и информативность.

1. ДСТУ ISO 19011:2003. Настанови щодо здійснення аудитів систем управління якістю і (або) екологічного управління (ISO 19011:2002, IDT). — К.: Держспоживстандарт України, 2004. — 24 с.

2. *Оптнер С. Л.* Системный анализ для решения проблем бизнеса и промышленности / Станфорд Л. Оптнер; пер. с англ. — М.: Концепт, 2006. — 206 с.
3. ДСТУ ISO 9001:2009. Системи управління якістю. Вимоги. Введ. 01.09.2009. — К.: Держспоживстандарт України, 2009. — 23 с.
4. *Машинная* геометрия и графика / Аджиев В.Д., Пасько А.А., Пилюгин В.В. и др. — М.: Знание, 1990. — 47 с.
5. *Факторный*, дискриминантный и кластерный анализ / Дж.-О. Ким, Ч.У. Мьюллер, У.Р. Клекка и др.; пер. с англ. под ред. И.С. Енюкова. — М.: Финансы и статистика, 1989. — 215 с.
6. *Енюков И.С.* Методы, алгоритмы, программы многомерного статистического анализа / И.С. Енюков. — М.: Финансы и статистика, 1989. — 232 с.
7. *Статистические* методы для ЭВМ / Под ред. К. Энслейна, Э. Рэлстона, Г.С. Уилфа; пер. с англ. под ред. М.Б. Малютова. — М.: Наука, 1986. — 464 с.

Поступила в редакцию 02.03.2010