

УДК 621.391.26

© Л. В. Новиков

ЭКСПРЕСС-АНАЛИЗ КАЧЕСТВА МНОГОКОМПОНЕНТНЫХ СМЕСЕЙ

С использованием имитационного моделирования исследован метод анализа качества многокомпонентных смесей (продуктов химической и пищевой промышленности) с использованием мультисенсорных систем. Предложена процедура отбраковки сенсоров с линейно-зависимой чувствительностью по рангу ковариационной матрицы откликов сенсоров и выбора их количества в системе по собственным числам этой матрицы. Моделируемая система осуществляет дискриминантный анализ (распознавание образов с учителем) многокомпонентных смесей по схеме: (обучение—измерение откликов сенсоров от контролируемого образца—вычисление вероятности их принадлежности обучающей выборке). Обработка данных ведется в пространстве главных компонент, что повышает вероятность правильной классификации.

Кл. сл.: многокомпонентный анализ, неселективные сенсоры, дискриминантный анализ, метод главных компонент

ВВЕДЕНИЕ

Одно из направлений развития новых методов анализа и аналитического приборостроения основано на применении в качестве чувствительных элементов металлооксидных, оптико-волоконных, электрохимических сенсоров, таких как ионоселективные электроды (ИСЭ), ионоселективные полевые транзисторы (ИСПТ) и др. [1, 2]. Созданные на их основе приборы привлекают невысокой стоимостью, малыми габаритами, имеют сравнительно высокую скорость экспресс-анализа растворов и газов. Однако многие современные химические и другие сенсоры при хорошей чувствительности имеют недостаточную селективность в многокомпонентных растворах, что сильно ограничивает их применение в аналитических целях. Выход был найден в создании мультисенсорных систем. Отклики сенсоров в таких системах образуют многомерный образ, который содержит информацию о качестве и количественном составе анализируемой смеси. Известны два способа извлечения этой информации из экспериментальных данных: решение уравнения Никольского и метод многомерной обработки [1, 3]. Для решения указанного уравнения требуется знание ряда параметров сенсоров, таких как коэффициенты селективности и активности основного и мешающего ионов, что сопряжено с рядом практических проблем. Поэтому предпочтительно использование второго подхода, не требующего априорной информации о сенсорах.

Сенсоры мультисенсорных систем должны реа-

гировать на все интересующие компоненты анализируемой смеси, т. е. обладать свойством перекрестной чувствительности. Они, кроме того, должны иметь воспроизводимые аналитические характеристики и неповторяемое от сенсора к сенсору покомпонентное распределение чувствительности.

В настоящей работе методом имитационного моделирования исследуется мультисенсорная система, в которой могут быть заданы число сенсоров, вид распределения чувствительности сенсора по компонентам (эпюра чувствительности), число компонентов анализируемой смеси и их концентрации. С помощью разработанной модели изучены возможности системы к распознаванию некачественных продуктов химической, фармацевтической или пищевой промышленности и количественной оценки несоответствия требуемому качеству. В основу обработки данных положен метод главных компонент (МГК), который широко применяется в задачах дискриминантного анализа и многомерной калибровки [4, 5].

ОПИСАНИЕ АЛГОРИТМА ИМИТАЦИОННОЙ МОДЕЛИ

Исследуемая мультисенсорная система показана на рис. 1 [3, 6]. Набор сенсоров погружен в многокомпонентный раствор. Разность потенциалов между каждым из сенсоров и электродом сравнения через коммутатор с предусилителями К поступает на вход аналого-цифрового преобразователя АЦП. Данные в цифровом виде обрабатываются

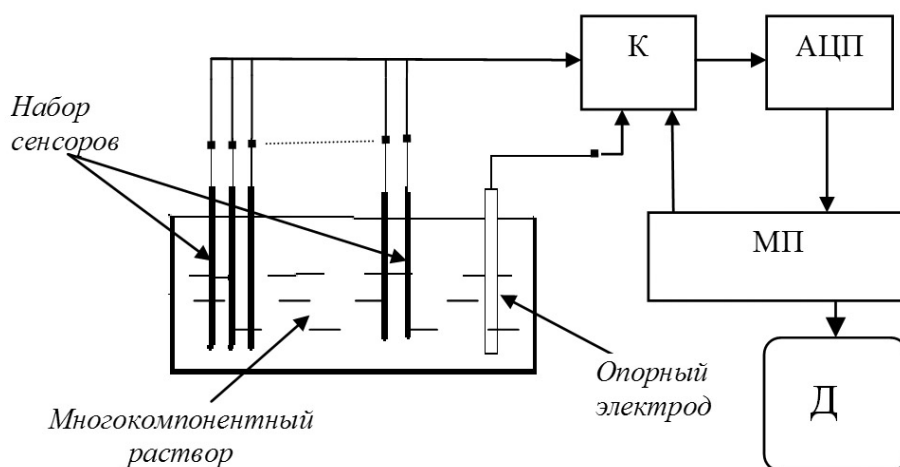


Рис. 1. Моделируемая мультисенсорная система

микропроцессором МП, который управляет также считыванием данных сенсоров. Результаты вычислений отображаются на дисплее Д.

Программа моделирования содержит следующие операции.

Подготовка обучающих образцов

1. Формирование эпюры распределения чувствительности по S компонентам для каждого из J сенсоров: $Sens(s, j)$, $s = 1, 2, \dots, S$; $j = 1, 2, \dots, J$.

2. Формирование набора I обучающих образцов многокомпонентной смеси, концентрации компонентов C в каждом из которых не превышает предельных уровней (ПДК): $C(k, s)$, $k = 1, 2, \dots, I$. Вычисление откликов сенсоров: $(I \times J)$ -матрицы XS , $XS = C \cdot Sens$ [4].

3. Проверка наличия линейной зависимости между эпюрами чувствительности сенсоров по рангу ковариационной матрицы $K = \begin{pmatrix} 0 \\ XS \end{pmatrix}^T \cdot \begin{pmatrix} 0 \\ XS \end{pmatrix}$,

где XS — центрированная по столбцам матрица XS , T — знак транспонирования.

Обучение

4. Вычисление собственных векторов (матрицы P) и собственных чисел $\lambda = [\lambda_1, \lambda_2, \dots, \lambda_J]$ $(J \times J)$ -ковариационной матрицы K и вычисление проекций обучающих образцов на оси ГК: $T = XS \cdot P$.

5. Оптимизация размерности пространства ГК J_0 и выбор числа сенсоров системы из условия $\lambda_{J_0} \ll \lambda_1$.

Анализ (классификация)

6. Формирование $(1 \times S)$ -вектора концентра-

ций компонент контрольной пробы CI , не превышающих ПДК.

7. Вычисление откликов сенсоров от контрольной пробы $XSI = CI \cdot Sens$.

8. Центрирование вектора XSI с помощью средних значений по столбцам матрицы XS .

9. Вычисление проекций контрольной пробы на оси главных компонент $T_c = XSI \cdot P$.

10. Вычисление величины вероятности соответствия контрольного образца требованиям ПДК.

11. Построение графических образов результатов вычислений в цикле.

12. Пошаговое изменение концентрации CI отдельных компонент контрольной пробы до уровня, превышающего ПДК.

13. Переход к п.7 и выход после завершения цикла.

14. Построение графических зависимостей по результатам вычислений.

Первый этап (пп. 1–3)

Это один из важнейших шагов подготовки мультисенсорной системы к работе. На этом этапе важно подобрать надлежащие образцы для обучения. При моделировании эти образцы формируются путем задания их концентраций C и эпюры чувствительности сенсоров $Sens$, после чего вычисляются их отклики по формуле п. 2 программы. В условиях производства это можно сделать после отладки технологии и проверки качества продукта в лабораторных условиях, выполнив I измерений при разных условиях (время суток, вид сырья и т. п.). Отклики сенсоров (их называют признаками) от всех образцов образуют матрицу объект—признак исходных данных $XS = [x_{i,j}]$,

где $i = 1, \dots, I$ — номер объекта (образца); $j = 1, \dots, J$ — номер признака. В J -мерном пространстве (пространстве измерений) отклики сенсоров выглядят, как "облако" из I точек обучающих образцов. Вследствие того что при измерении параметров идентичных объектов сенсорами с перекрестной чувствительностью неизбежно наличие корреляционных связей между откликами сенсоров, "облако" вытянуто в пространстве, т. е. имеет вид веретена [4, 5]. При проверке принадлежности контролируемого образца одному из классов в этом же пространстве может быть сформировано несколько "облаков".

На этапе подготовки целесообразно также проверить наличие сенсоров с линейной зависимыми эпюрами чувствительности. Для этого необходимо вычислить ранг $(J \times J)$ ковариационной матрицы $\mathbf{K}$. Если этот ранг равен J , то линейная зависимость отсутствует. В противном случае нужно удалить (или заменить) один или несколько сенсоров, связанных друг с другом линейной зависимостью.

Этап обучения (пп. 4, 5)

На этапе обучения формируется пространство главных компонент (ГК). Для получения статистически устойчивых результатов число образцов I должно быть большим (по крайней мере больше десяти).

Базисные векторы пространства ГК $\mathbf{p}_r = [p_{1,r}, p_{2,r}, \dots, p_{I,r}]$, $(r = 1, \dots, J)$ образуют векторы-столбцы матрицы $\mathbf{P}$, которую называют матрицей нагрузок (loadings). Представление исходных данных $\mathbf{XS}$ в новой системе координат производится по формуле п. 4.

Матрицу $\mathbf{T}$ называют матрицей счетов (scores). Столбцы матрицы $\mathbf{T}$ обозначают $\mathbf{t}_j = [t_{1,j}, t_{2,j}, \dots, t_{I,j}]$, $(j = 1, \dots, J)$ и называют главными компонентами, а $t_{1,j}, t_{2,j}, \dots, t_{I,j}$ — проекции первого, второго и т. д. образцов на j -ю координатную ось. Все эти проекции являются элементами множества, которое для дальнейшего обозначим как x_j ($x \in R, j = 1, 2, \dots, J$).

Вычислением матриц $\mathbf{P}$ и $\mathbf{T}$ завершается процесс обучения системы.

Существенной важнейшей для практики особенностью МГК является способность максимизировать дисперсию первых ГК, что позволяет без потери информации удалить часть последних столбцов матрицы $\mathbf{T}$, сократив ее до размера $I \times J_0$ ($J_0 < J$), т. е. уменьшить размерность пространства ГК и соответственно число сенсоров до

величины J_0 . Для этой цели используется критерий Кайзера, или критерий "каменистой осыпи", по которому собственное число (дисперсия) самой старшей J_0 -й ГК $\lambda_{J_0} \ll \lambda_1$ [4].

Еще одной особенностью МГК является некоррелированность столбцов $\mathbf{t}_j$ матрицы $\mathbf{T}$, т. е. ковариация от $\mathbf{T}$ образует диагональную матрицу со значениями дисперсии в диагонали $\sigma_{j,j}^2 = \sigma_j^2 = (\mathbf{t}_j - \bar{\mathbf{t}}_j)^T (\mathbf{t}_j - \bar{\mathbf{t}}_j)$, $j = (1, \dots, J_0)$, где $\bar{\mathbf{t}}_j$ — среднее значение по j -му столбцу матрицы $\mathbf{T}$, причем $\sigma_1 > \sigma_2 > \dots > \sigma_{J_0}$. Так как $\bar{\mathbf{t}}_j = 0$ для всех j , то $\sigma_{j,j}^2 = \sigma_j^2 = \mathbf{t}_j^T \mathbf{t}_j$.

Этап анализа

На этапе анализа (или классификации) решается задача отнесения контролируемой пробы к продуктам соответствующего качества, или к какому-либо классу (пп. 6–13). Если, например, точка от контролируемой пробы в J_0 -мерном пространстве находится внутри "облака", то контролируемый образец соответствует требуемому качеству. Поэтому очень важно зарегистрировать ситуацию, когда точка от этой пробы находится вне "облака". Это можно легко сделать в пространстве ГК, т. к. в этом случае координатные оси МГК упорядочены по убыванию дисперсии проекций исходных данных на эти оси: первая главная компонента (ГК1) проходит по оси веретена, вторая — перпендикулярно ей вдоль максимальной дисперсии и т. д. В процессе анализа по формуле п. 9 вычисляется вектор $\mathbf{T}_c = (t_{c,1}, t_{c,2}, \dots, t_{c,J_0})$ проекций на оси ГК откликов сенсоров от контролируемого образца.

За количественную оценку качества контролируемого образца примем вероятность его попадания в "облако". Для ее вычисления используем J_0 главных компонент обучающих образцов (матрицу $I \times J_0$) и построим функцию многомерной плотности вероятности, предполагая закон распределения нормальным. Учитывая отмеченные выше свойства МГК, эта функция может быть записана в виде

$$w(x_1, \dots, x_{J_0}) = \frac{1}{(2\pi)^{J_0/2} \sigma_1 \sigma_2 \dots \sigma_{J_0}} \exp \left\{ -\sum_{k=1}^{J_0} \frac{x_k^2}{2\sigma_k^2} \right\}. \quad (1)$$

За количественную оценку степени близости контролируемого образца к совокупности обучающих (т. е. надлежащего качества) образцов можно принять нормированную вероятность

$$P_c = (2\pi)^{J_0/2} \sigma_1 \sigma_2 \dots \sigma_{J_0} \int_{t_{c,1}-\delta/2}^{t_{c,1}+\delta/2} \dots \int_{t_{c,J_0}-\delta/2}^{t_{c,J_0}+\delta/2} w(x_1, \dots, x_{J_0}) dx_1 \dots dx_{J_0} = \exp \left\{ -\sum_{k=1}^{J_0} \frac{t_{c,k}^2}{2\sigma_k^2} \right\}, \quad (2)$$

где $t_{c,k}$ — проекция контролируемого образца на k -ю ось и интегрирование плотности вероятности (1) ведется в δ -окрестности точек $t_{c,k}$.

Нетрудно убедиться в том, что выражение под знаком $\exp$ в формуле (2) определяет относительное евклидово расстояние между точкой контролируемого образца в J_0 -мерном пространстве и центром тяжести, центроидом, "облака", координатами которого являются средние значения $\bar{t}_k = 0$, $k=1, 2, \dots, J_0$. Обозначим это расстояние как

$$d_c, \text{ т. е. } d_c = \left\{ \sum_{k=1}^{J_0} \frac{t_{c,k}^2}{\sigma_k^2} \right\}^{1/2}. \text{ Тогда}$$

$$P_c = \exp \left\{ -\frac{1}{2} d_c^2 \right\}. \quad (3)$$

Формулы (1)–(3) используются в п. 10 программы моделирования.

Величина P_c является обобщенной характеристикой качества контролируемого продукта. Непрерывный мониторинг P_c на технологической линии дает возможность оперативно отреагировать на ухудшение этого качества. Во всех случаях из практических соображений должны быть установлены некоторые границы качества, например, при $1 \geq P_c > 0.6$ — продукт хорошего, при $0.6 \geq P_c > 0.2$ — среднего, при $P_c \leq 0.2$ — пограничного качества.

РЕЗУЛЬТАТЫ МОДЕЛИРОВАНИЯ

Рассмотрим мультисенсорную систему, сенсоры которой чувствительны в разной степени ко всем компонентам анализируемой смеси, число которых принято равным $J=12$. Пример эпюры чувствительности трех сенсоров приведен на рис. 2. Очевидно, что в этом случае изменение концентрации любой компоненты будет отражено в откликах всех сенсоров набора, что затрудняет или делает невозможным идентифицировать компоненту вещества, по которой произошло превышение ПДК. Тем не менее превышение контрольного образца над "облаком" обучающих образцов может быть обнаружено в процессе обработки откликов сенсоров.

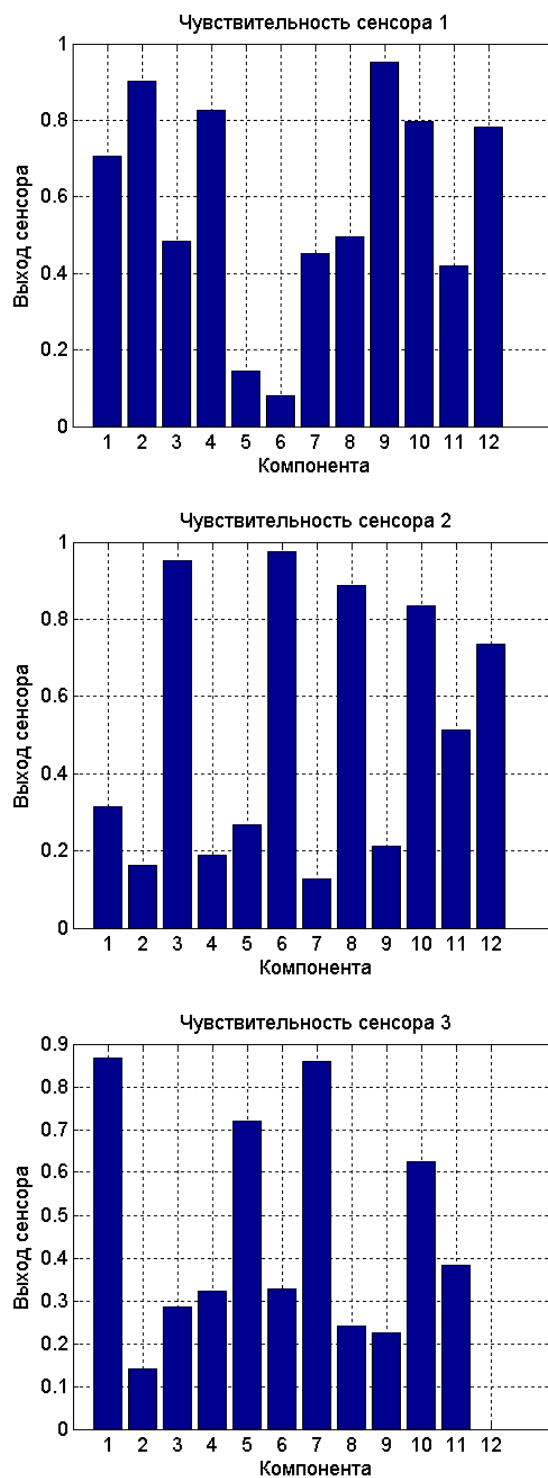


Рис. 2. Примеры покомпонентной чувствительности неселективных сенсоров

На рис. 3 показано на примере двумерного пространства, что для решения этой задачи пространство главных компонент (ГК1, ГК2) имеет преимущество перед пространством измерений (X_1 , X_2), т. к. проекция контролируемого

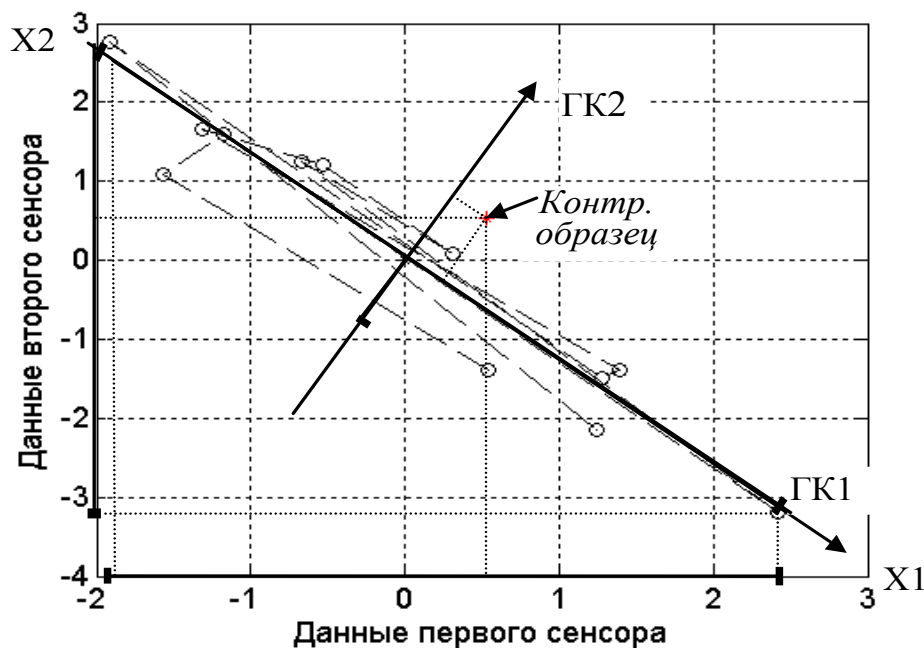


Рис. 3. "Облако" обучающих образцов (о) и точка контрольного образца в двумерном пространстве. X_1 , X_2 — координатные оси пространства измерений; GK_1 , GK_2 — оси пространства главных компонент

образца на ось GK_2 в отличие от других проекций удалена от зоны рассеяния точек обучающих образцов. Это обстоятельство позволяет с большой вероятностью сделать вывод о несоответствии контролируемого образца требуемому качеству.

Необходимое количество сенсоров в мультисенсорной системе можно оценить по графику изменения собственных чисел $\lambda = [\lambda_1, \lambda_2, \dots, \lambda_J]$ (вычисляемых в п. 4) от номера ГК (рис. 4), из которого следует, что для выбранного набора сенсоров и обучающих образцов размерность пространства (и число сенсоров) может быть сокращена с $J=12$ до $J_0=6$.

Выполним далее вычисления по пп. 8–13 и п. 14, приняв для определенности границы допустимых концентраций равными среднеквадратическому отклонению σ_k , $k=1, 2, \dots, J_0$, по каждой из координатных осей ГК. На рис. 5, а, показан пример, когда контролируемый образец соответствует требуемому качеству (находится в зоне ПДК), а на рис. 5, б, — одна из компонент превышает ПДК, причем по одной из координат (GK_5) это превышение максимально. Если другие компоненты вещества превышают ПДК, то распределения проекций по осям ГК будут отличаться. Этим обстоятельством можно воспользоваться для идентификации компонент. На рис. 6 показаны проекции (отклики сенсоров при превышении ПДК) на координатные оси пространства измерений. Сравнив этот рисунок с рис. 5, б, можно заметить, что в пространстве ГК уровни контроли-

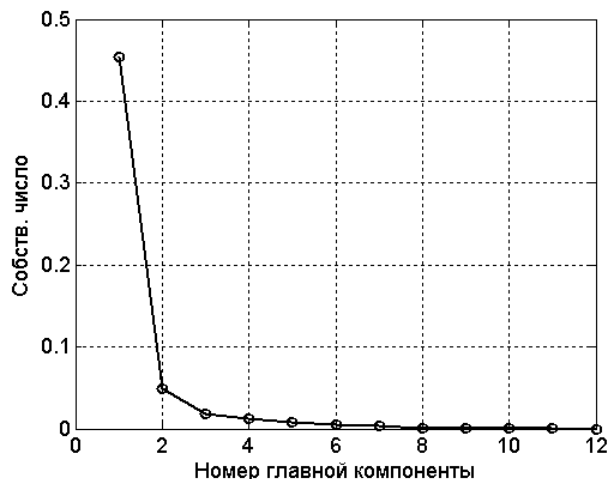


Рис. 4. График распределения собственных чисел ковариационной матрицы данных обучающих образцов K по главным компонентам

руемого образца больше отдалены от зоны ПДК, чем в пространстве измерений (см. также рис. 3).

На рис. 7 показана зависимость вероятности (2) нахождения контрольного образца в "облаке" обучающих образцов: эта вероятность, естественно, уменьшается (и становится меньше 0.6) после достижения одной из проекций (в данном случае GK_5 (рис. 5, б)) верхнего уровня ПДК ($d_c=1$ по оси абсцисс). Это означает, что если полученная

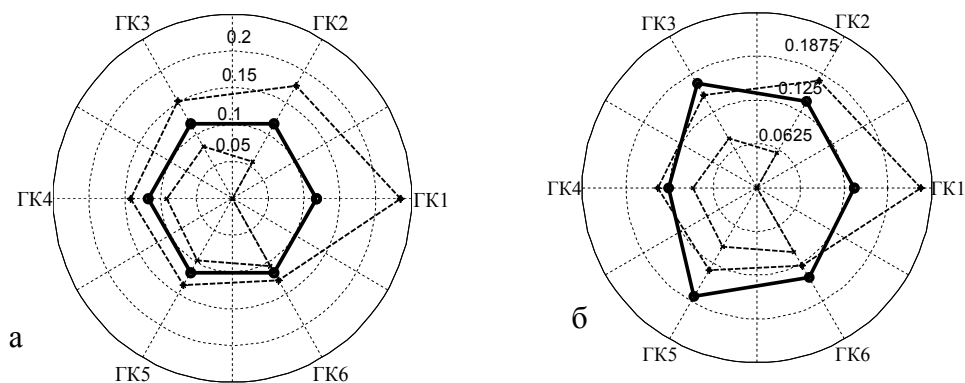


Рис. 5. Уровни главных компонент контрольного образца в полярных координатах.

а — концентрации в пределах ПДК; б — одна из компонент превышает ПДК (штриховые линии — нижний и верхний уровни ПДК)

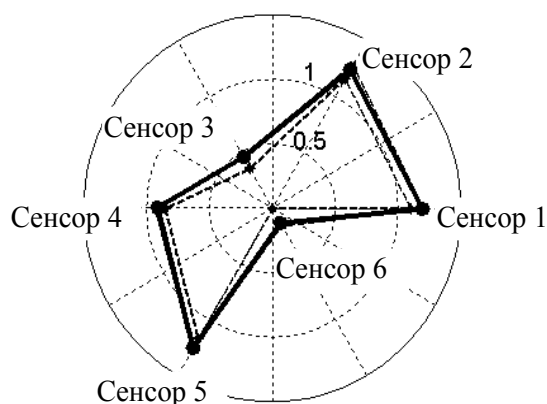


Рис. 6. Отклики сенсоров при превышении одной из компонент ПДК



Рис. 7. Вероятность классификации контрольного образца

по формуле (3) вероятность будет больше 0.6 (для принятой границы допустимых концентраций), то этот образец соответствует требуемому качеству с вычисленной вероятностью и т. д.

ВЫВОДЫ

В работе предложены подходы для выбора числа сенсоров в мультисенсорных системах количественной оценки качества продуктов, как вероятности классификации. Показано преимущество использования метода главных компонент для решения этих задач.

На основе неселективных сенсоров с перекрестной чувствительностью может быть создан недорогой прибор для экспресс-анализа качества многокомпонентных смесей. Прибор гарантирует высокую вероятность распознавания некачественного продукта, в частности, химической или пищевой промышленности.

СПИСОК ЛИТЕРАТУРЫ

1. Легин А.В., Рудницкая А.М., Власов Ю.Г. "Электронный язык" системы химических сенсоров для анализа водных сред // Проблемы аналитической химии. 2011. Т. 14. С. 72–119.

2. Щербакова Л.Ф., Шантроха А.В., Егоров И.В., Щербаков А.А. Сенсорные системы для экспресс-определения токсичных химикатов // Российский химический журнал. 2007. Т. LI, № 2. С. 127–131.
3. Власов Ю.Г., Легин А.В., Рудницкая А.М. Мультисенсорная система типа электронный язык — новые возможности создания и применения химических сенсоров // Успехи химии. 2006. Т. 75, № 2. С. 141–150.
4. Большаков А.А., Каримов Р.Н. Методы обработки многомерных данных и временных рядов. М.: Горячая линия-Телеком, 2007. 522 с.
5. Эсбенсен К. Анализ многомерных данных. Черноголовка: Изд-во ИПХФ, 2005. 158 с.
6. Кисляков Ю.Я., Кислякова Л.П., Зайцева А.Ю. Обу-

чаемая мультисенсорная электрохимическая система "электронный язык" для контроля биологических сред // Материалы XVI Международной конференции по нейрокибернетике, 24–28 сентября 2012 г., Ростов-на-Дону. С. 93–96.

*Институт аналитического приборостроения РАН,
г. Санкт-Петербург*

Контакты: Новиков Лев Васильевич,
novik38@mail.ru

Материал поступил в редакцию 20.01.2014

EXPRESS QUALITY ASSAY OF MULTICOMPONENT MIXTURE

L. V. Novikov

Institute for Analytical Instrumentation of RAS, Saint-Petersburg, RF

Using simulation modeling we analyzed the method of quality assay of multicomponent mixture (products of chemical and food industries) that applies multisensory systems. We suggested the procedure of sensor rejection with linearly dependent sensibility according to matrix rank of correlation matrix of sensor response. We also proposed the selection of sensors in the system following the eigenvalues of this matrix. The simulated system realizes the discriminatory analysis (supervised pattern recognition) of multicomponent mixtures according to the scheme: (supervision—measuring of sensor responses from the controlled sample—calculation of the probability of their belonging training sample). Data processing is realized in the space of principal components, that increases the probability of correct classification.

Keywords: multi-component analysis, nonselective sensor, discriminatory analysis, principal component analysis

Contacts: Novikov Lev Vasilyevich,
novik38@mail.ru

Article arrived in edition: 20.01.2014