УДК 664:004.42:543.92

Методологическое обеспечение и разработка программы обработки сенсорных данных, полученных с помощью профильно-дескрипторного анализа

М.А. Никитина, канд. техн. наук; **Т.Г. Кузнецова**, д-р вет наук;
А.А. Лазарев, канд. техн. наук; **А.Н. Захаров**, канд. техн. наук
Федеральный научный центр пищевых систем имени В.М. Горбатова РАН, Москва

Одним из приоритетных направлений стратегической политики по развитию предприятий-производителей пищевой продукции является мониторинг и анализ запросов покупателей в отношении ее качества на потребительском рынке. Поскольку органолептические свойства относятся к характеристикам продукта, непосредственно влияющим на потребительский спрос, производители заинтересованы в их изучении с помощью современных методов сенсорного анализа [1].

В последнее время все чаще производители обращаются к комплексным, сложным, аналитическим методам, таким как профильно-дескрипторный (профильный) анализ, который позволяет не только описать продукт во всех воспринимаемых ощущениях – зрительных, слуховых, обонятельных, кинестетических, но и определить величину сенсорных различий данных ощущений [2]. Своеобразные подходы и философия этого анализа дает возможность получить универсальные решения для большинства задач, стоящих перед дегустатором. Так, с помощью профильно-дескрипторного анализа можно разрабатывать новые продукты и совершенствовать сенсорные свойства производимых продуктов, изучать стабильность сенсорных свойств при хранении, устанавливать сенсорные спецификации, проводить мониторинг свойств продуктов-конкурентов, прогнозировать изменения и определять слабые стороны продукта и многое другое [3, 4].

Применение профильно-дескрипторного анализа требует использования сложных методов статистической обработки данных, которые позволяют утверждать, что полученные результаты являются объективными и достоверными.

К основным статистическим методам, используемым для обработки данных, полученных методами профилирования, относятся методы многомерной статистики [5]. Существует достаточно много методов многомерного анализа (в том числе – дисперсионный, дискриминантный или факторный анализ) и прежде чем использовать эти методы, следует определить наиболее подходящий в каждом конкретном случае.

К сожалению, на российском рынке программного обеспечения предприятий пищевой промышленности, в отличие от зарубежного, практически отсутствуют специализированные программы для анализа сенсорных данных (в т.ч. профильно-дескрипторного анализа). Оперативные мобильные программы, с возможностями сбора, накопления, обработки полученных (сенсорных) данных, представлении их в удобной форме на сегодняшний день являются перспективой будущего для российских производителей. В связи с этим специалистами ФНЦ пищевых систем им. В.М. Горбатова РАН была поставлена задача: разработать методологию создания программного обеспечения для обработки сенсорных данных на основе анализа современных статистических методов.

Большинство статистических методов, используемых для обработки сенсорных данных (в том числе, полученных в результате дегустации), направлены на поиск значимых различий в оценках образцов (экспертами, потребителями) и /или наличие значимых связей между сенсорными характеристиками (переменными). При этом формат выходных данных (обусловленный типом шкалы, количеством дегустаторов и др. факторами) требует соответствующих методов обработки. С методологической точки зрения для исследования связей применяют параметрические или непараметрические корреляции и регрессию, а в случае исследования различий – дисперсионный анализ, спектр непараметрических методов и апостериорных тестов [6].

Обычно для простых сенсорных исследований используют однофакторные методы, обеспечивающие надежные и достоверные результаты. Тем не менее, данные, имеющие многоуровневую структуру, анализируют многофакторными статистическими процедурами [7]. Так, многофакторный дисперсионный анализ позволяет определять различия между двумя или более наборами данных по всем зависимым переменным одновременно. Это способствует сохранению низкого уровня общей ошибки первого рода и решению вопроса взаимодействия между зависимыми переменными. Кроме того, многофакторный дисперсионный анализ помогает установить комбинации сенсорных переменных, позволяющих различать образцы, в случае, когда не обнаруживается отличие по каждой из них в отдельности [8].

Другим, актуальным многомерным методом, является метод главных компонент, который упрощает и описывает взаимосвязь между несколькими зависимыми переменными (как правило, дескрипторами) и объектами (как правило, продуктами) [9]. Принцип работы метода заключается в преобразовании исходных зависимых переменных в новые некоррелируемые размерности, что особенно важно в сенсорном анализе, где несколько зависимых переменных часто являются коллинеарными. Например, в описательных исследованиях, эксперты часто оценивают запах и флейвор в продукте, при этом возможно, что такой подход является избыточным, поскольку и запах, и флейвор измеряют одну характеристику [10]. Метод главных компонент позволяет выявлять избыточности путем преобразования данных в новый набор переменных, называемый основными компонентами.

Многофакторный дисперсионный анализ и метод главных компонент являются наиболее распространенными многомерными методами для анализа данных большинства классических описательных методов. Тем не менее, в сенсорном анализе используют и другие процедуры обработки данных. Например, дискриминантный и канонический

вариационный анализ используются в классификации образцов продукции и анализе взаимосвязей между их дескрипторами [11, 12]. При этом они позволяют получить более достоверные результаты, чем метод главных компонент. Дискриминантный и канонический вариационный анализ относятся к многомерной технике значимого разделения, аналогично многофакторному дисперсионному анализу. Эти методы особенно полезны, когда нужно использовать исходные данные, чтобы получить информацию о межпродуктовой или внутрипродуктовой изменчивости.

Общий прокрустов анализ можно использовать для сравнения гедонических и описательных сенсорных данных; описательной информации, полученной от разных экспертных групп различными методами; результатов, инструментальных и сенсорных исследований; проблемах согласованности используемых терминов [13, 14]. При выполнении общего прокрустова анализа данные каждого дегустатора подвергаются смещению в фазе стандартизации, а оценки центрируются по их происхождению (исходным данным). Преимуществом общего прокрустова анализа является минимизация различий в терминах, использованных каждым дегустатором. Это дает возможность учитывать применение различных терминов для описания одного и того же ощущения.

Выбор метода статистической обработки для реализации задачи создания программного обеспечения по обработке сенсорных данных основывался на его возможности устанавливать изменчивость и различия между оценками дегустаторов.

Применение однофакторного дисперсионного анализа в данном случае является недостаточным (некорректным), поскольку требуется анализировать несколько объектов с множеством переменных. Такую позицию разделяет ряд авторов, установивших, что использование однофакторного дисперсионного анализа в случае обработки результатов оценки нескольких продуктов по нескольким дескрипторам дает завышение ошибки I типа и невозможность точной оценки ее увеличения [15]. Таким образом, в качестве базового статистического метода был предложен многофакторный (двухфакторный) дисперсионный анализ. Необходимо отметить, что аналогичные подходы по статистической обработке данных рекомендуются международными стандартами ISO в области сенсорного анализа [16, 17].

Таблица 1

Дисперсионная таблица для двух факторов (x_1 и x_2)

Уровень X_1	Уровень X_2				$\bar{y}_i = \frac{1}{n_2 q}\sum_{j=1}^{n_2}\sum_{k=1}^{q} y_{ijk}$
	1	2	...	n_2	
1	$Y_{1,1,1}$ $Y_{1,1,2}$ ... $Y_{1,1,q}$	$Y_{1,2,1}$ $Y_{1,2,2}$ ... $Y_{1,2,q}$	... ...	$Y_{1,n2,1}$ $Y_{1,n2,2}$ ... $Y_{1,n2,q}$	
...			...		...
n_1	$Y_{n1,1,1}$ $Y_{n1,1,2}$ ... $Y_{n1,1,q}$	$Y_{n1,2,1}$ $Y_{n1,2,2}$ ... $Y_{n1,2,q}$	... ...	$Y_{n1,n2,1}$ $Y_{n1,n2,2}$ ... $Y_{n1,n2,q}$	
$\bar{y}_j = \frac{1}{n_1 q}\sum_{i=1}^{n_1}\sum_{k=1}^{q} y_{ijk}$	$\bar{y}_1$	$\bar{y}_2$		$\bar{y}_{n_1}$	$\bar{y} = \frac{1}{n_1 n_2 m}\sum_{i=1}^{n_1}\sum_{j=1}^{n_2}\sum_{k=1}^{q} y_{ijk}$

Идея дисперсионного анализа состоит в выявлении влияния того или иного изучаемого фактора на отклик путем сравнения факторной дисперсии ($S^2_{факт}$), обусловленной воздействием исследуемого фактора, с остаточной дисперсией ($S^2_{ост}$), учитывающей действие случайных неучтенных факторов [18, 19].

Для того чтобы можно было использовать дисперсионный анализ, данные должны представлять собой набор выборок с одной и той же базовой единицей измерения. Необходимо обеспечить, чтобы данные удовлетворяли основным допущениям. Во-первых, предполагается, что каждая выборка представляет собой случайную выборку из генеральной совокупности, для которой будет выполняться обобщение. Во-вторых, предполагается, что каждая генеральная совокупность подчиняется нормальному распределению и стандартные отклонения этих совокупностей равны между собой.

Таким образом, применение двухфакторного дисперсионного анализа ограничено следующими требованиями:

– анализируемые выборки должны быть отобраны из генеральных совокупностей случайным образом;

– выборки должны быть независимыми и все оценки в выборке также должны быть независимыми (т. е. один испытатель может дать только одну оценку в каждой группе);

– данные должны иметь нормальное распределение, т. е. для оценки должны использоваться интегральные шкалы или шкалы соотношений;

– дисперсии генеральных совокупностей, из которых берутся выборки, должны быть одинаковы (гомосцедастичны).

Аддитивная математическая модель двухфакторного дисперсионного анализа может быть представлена в виде формулы 1:

$$X = \mu + \beta + S + \varepsilon \qquad (1)$$

где X – отклик (оценка); μ – среднее значение; β – отклонение от среднего; S – отклонение от среднего обусловленного индивидуальными различиями между субъектами; ε – случайная ошибка.

Параметрический двухфакторный дисперсионный анализ основан на предположении о нормальном распределении и гомосцедастичности экспериментальных данных. Для него выполняется соотношение $F=t^2$, в случае сравнения пары зависимых выборок. В этом случае дисперсионный анализ предпочтительнее, поскольку позволяет оценить не только различие между выборками, но и между субъектами, проводящими оценку.

При двухфакторном дисперсионном анализе используется группировка по двум факторам, то есть дополнительно исследуются индивидуальные различия между дегустаторами в группе.

В дисперсионном анализе оценивается межгрупповая и внутригрупповая дисперсия, которая при двухфакторном анализе включает в себя, помимо экспериментальной ошибки, еще и дисперсию, обусловленную индивидуальными различиями дегустаторов.

Дисперсионная таблица для случая двух факторов приведена в табл. 1.

Фактор x_1 варьирует на n_1 уровнях, а фактор x_2 варьирует на n_2 уровнях. При каждом сочетании уровней факторов проводится серия из q параллельных опытов, т. е. результаты эксперимента $n=n_1 \cdot n_2 \cdot q$ наблюдений y_{ijk}, где i – порядковый номер уровня фактора x_1; j – порядковый номер уровня фактора x_2; k – порядковый

номер параллельного опыта при каждом ij-м сочетании уровней факторов. Число параллельных опытов в каждой серии – постоянная величина (q=const).

Обработка данных осуществляется по определенному алгоритму, укрупненная блок-схема которого представлена на рис. 1.

На первом этапе происходит расчет средних значений по строкам, столбцам и всех значений (общего среднего).

Для оценки степени влияния тех или иных факторов на результаты эксперимента (2 этап), общую сумму квадратов отклонений наблюдений от общего среднего раскладывают на компоненты, зависящие от рассматриваемых факторов (фактор случайности, фактор взаимодействия, факторы x_1 и x_2).

Рис. 1. Укрупненная блок схема двухфакторного дисперсионного анализа

На третьем этапе производят оценку дисперсий. Каждая из полученных сумм, разделенная на соответствующее ей число степеней свободы, дает несмешанную оценку соответствующей дисперсии.

Анализ существенности влияния факторов и их взаимодействия (4 этап) производится с использованием критерия Фишера-Снедекора при выбранном уровне значимости.

Разработанное программное обеспечение на основе математических и статистических моделей обеспечивает решение задач по сбору и статистической обработке сенсорных данных [20]. Функциональная структура системы состоит из 6 модулей:

- параметры оценки;
- оценка дескрипторов продукта;
- создание набора данных для анализа;
- сенсорный профиль;
- сравнение с эталоном;
- помощь (пользователю и администратору).

В начале работы администратор (ПО) задает настройки параметров оценки и проведения дегустации (рис. 2). Перечень параметров, определяемый целью дегустации, включает в себя:

- количество дегустируемых образцов и оцениваемых дескрипторов;
- тип шкалы (структурированная или неструктурированная);
- наименование оцениваемых дескрипторов;
- папка для сохранения файлов в формате *.txt с результатами дегустации;
- инструкция для дегустационной комиссии.

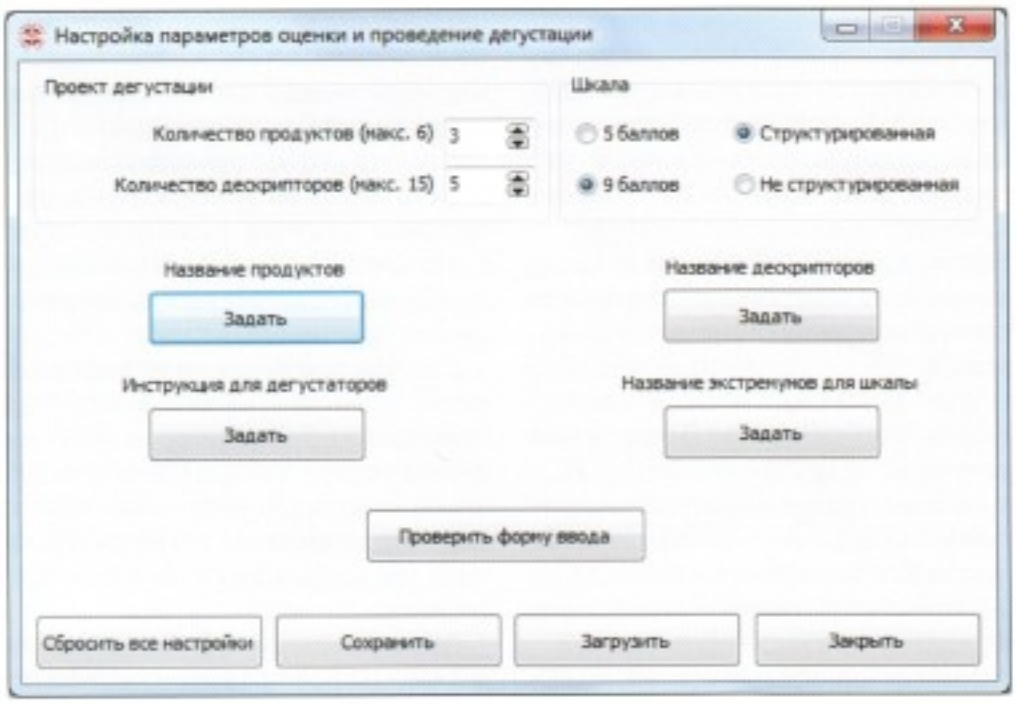

Рис. 2. Модуль параметров настройки

После установки параметров оценки проводится подключение дегустаторов к Программе-Сервер. Задачи дегустатора максимально ограничены в плане работы с программой. Он просто вводит идентификационные данные (например, Ф.И. О.) в Программе-Клиент и нажимают кнопку «Приступить» (рис. 3). Далее дегустатор оценивает с помощью шкалы интенсивность дескрипторов в образцах продукции. Для непосредственной оценки дескриптора он передвигает курсор (бегунок) по линии шкалы от отметки «минимального экстремума» до отметки «максимального экстремума». После оценки всех дескрипторов в продукте дегустатор, нажатием кнопки, переходит к оценке следующего продукта или завершает дегустацию (в случае, если оценивается 1 продукт).

При органолептической оценке образцов дегустаторы (в Программе-Клиент) могут регистрировать характеристики (дескрипторы) продукта, порядок их проявления, интенсивность и общее впечатление. После окончания дегустации результаты оценочных операций передаются на компьютер-сервер (Программа-Сервер), где производится статистическая обработка индивидуальных оценок, рассчитываются комплексные показатели и количественные меры согласованности дегустаторов.

Модуль «Сенсорный профиль» (рис. 4) позволяет загружать ранее сохраненный массив данных (по критериям – образец, дегустатор, дескриптор); визуально отображать количество выбранных объектов в массиве (образцов, дегустаторов, дескрипторов); удалять массив данных; проверять общую согласованность всей дегустационной комиссии по всем дескрипторам всех образцов; визуально отображать результаты проверки общей согласованности по объектам словами «Согласовано» или «Не согласовано»; сохранять результаты проверки общей согласованности по объектам; графически отображать результаты несогласованной оценки дегустатора по дескрипторам и образцам; сохранять результаты несогласованной оценки дегустатора в графическом и числовом формате; строить сенсорные профили на общем графике по образцам продукции в случае согласованной оценки; добавлять и удалять

Рис. 3. Подключение дегустатора к программе Программе-Сервер через Программу-Клиент

профили образцов на общем графике; сохранять результаты построения сенсорных профилей в графическом и числовом формате (в том числе в качестве эталона).

Детальный анализ, почему тот или иной дескриптор согласован или нет, можно посмотреть нажав на соответствующую данной характеристике оценку. На рис. 5 представлена таблица результатов двухфакторного дисперсионного анализа по дескриптору – «Запах копчения».

Как видно из рис. 5, расчетное значение F-критерия фактора x_1 (продукты) F=19,85, а критическая область образуется правосторонним интервалом (4,46;+∞). Так как F попадает в критическую область, то нулевая гипотеза отвергается, т.е. считаем, что фактор x_1 (продукты) влияет на дескриптор.

Расчетное значение F-критерия фактора x_2 (дегустаторы) F=1, а критическая область образуется правосторонним интервалом (3,84; +∞). Так как F не попадает в критическую область, то нулевая гипотеза принимается, т.е. считаем, что влияние фактора x_2 (дегустаторы) на дескриптор не подтвердилось.

Выборочный коэффициент детерминации

$$\bar{\rho}^2_{x1} = \frac{SS_{x1}}{SS_{x1}+SS_{x2}+SS_e} = \frac{17.2}{17.2+1.73+3.47} \approx 0.77$$

показывает, что 77 % общей выборочной вариации качества дескриптора (запах копчения) связано с влиянием на него вида продукции.

В столбце P-значение определено P-значение, соответствующее расчетному значению критерия F.

Функция F-распределения (распределение Фишера, называемое иногда распределение дисперсионного отношения). Распределение Фишера – случайная величина, равная отношению двух независимых случайных величин: величины $\chi^2(v_1)/v_1$ с распределением χ^2 и v_1 степенями свободы и величины $\chi^2(v_2)/v_2$ с распределением χ^2 и v_2 степенями свободы.

$$F(v_1, v_2) = \frac{\chi^2(v_1)}{v_1} \cdot \frac{v_2}{\chi^2(v_2)}.$$

В нашем случае, P-значение для фактора x_1 (продукты) зависит от значений F, df и MS данного фактора, находящихся в первой строчки таблицы. P-значение для фактора x_2 (дегустаторы) зависит от значений F, df и MS данного фактора, находящихся вовторой строчки таблицы.

Если P-значение меньше 0,05 (P<0,05), то данные не согласуются.

В нашем случае для фактора x_1 (продукты) P=0,00079, для фактора x_2 (дегустаторы) P=0,46. Таким образом, делают вывод «Продукты различаются по данному дескриптору, оценки дегустаторов согласованны».

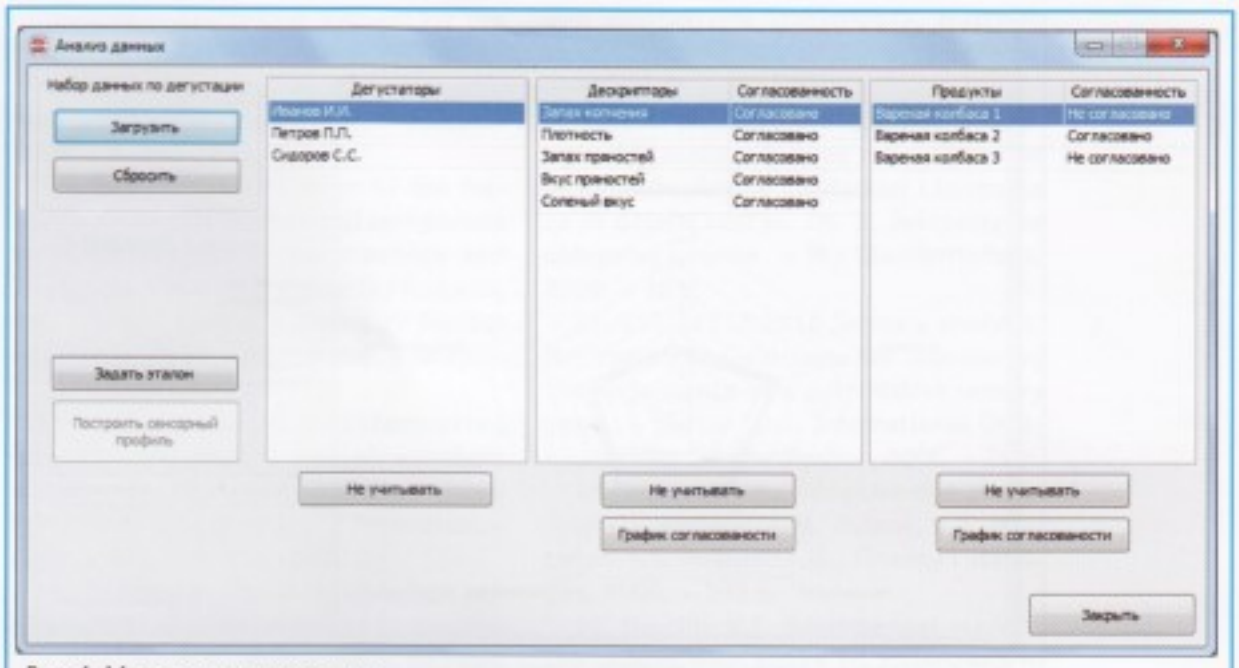

Рис. 4. Модуль анализа данных

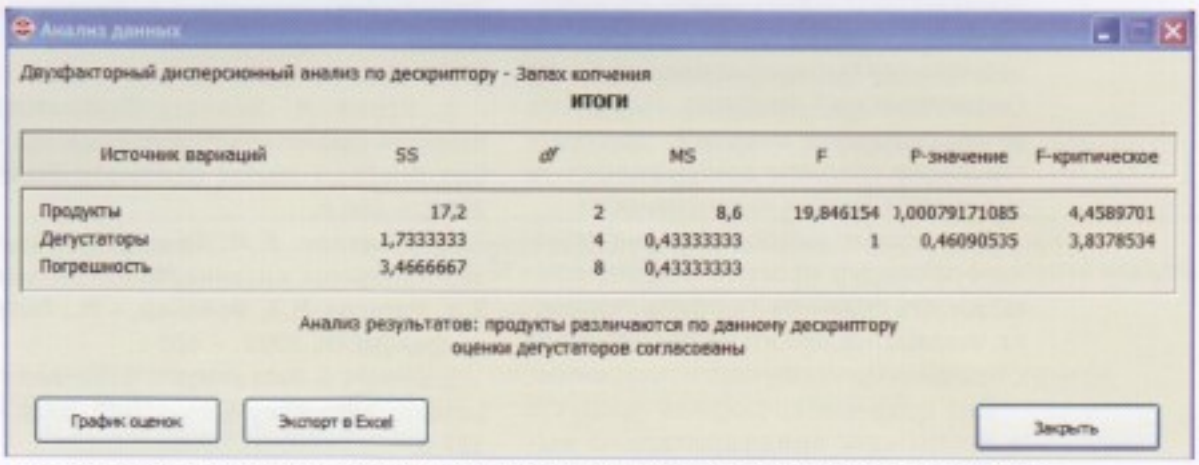

Рис. 5. Результаты проверки согласованности дегустаторов по дескриптору «Запах копчения»
SS – дисперсия; df – степень свободы; MS – несмещенные оценки; F – расчетный критерий Фишера; P-значение – функция F-распределения; F-критическое – табличное значение критерия Фишера

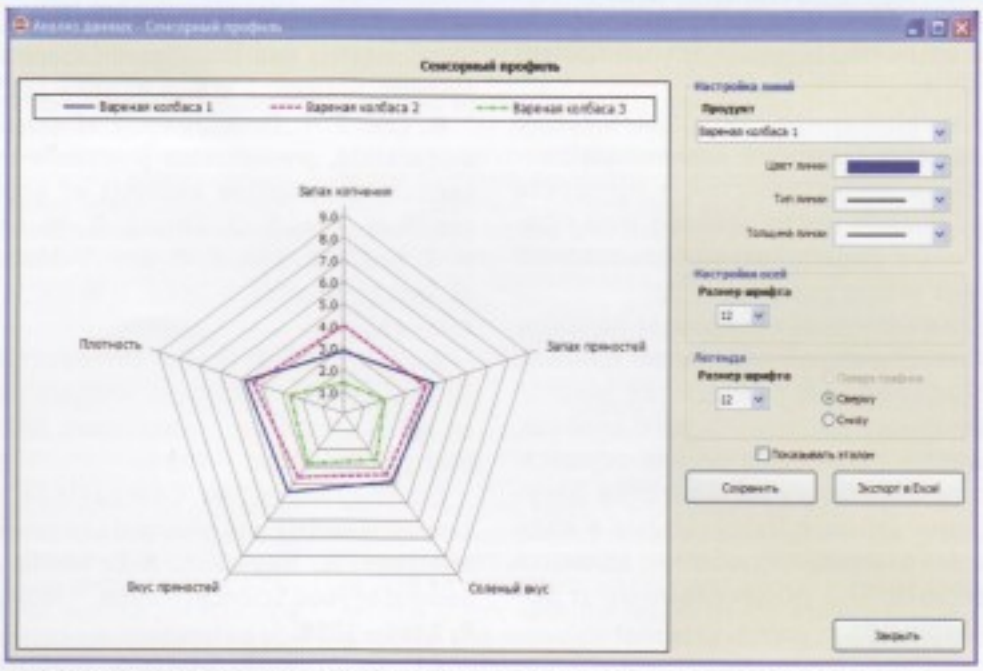

Рис. 6. Сенсорный профиль трех образцов вареной колбасы в Программе-Сервере

После статистической обработки результатов дегустации (в случае согласованной и достоверной оценки) программа позволяет построить сенсорные профили продуктов. На рис. 6 показан пример сенсорного профиля трех образцов вареной колбасы. Радиальные линии на графике

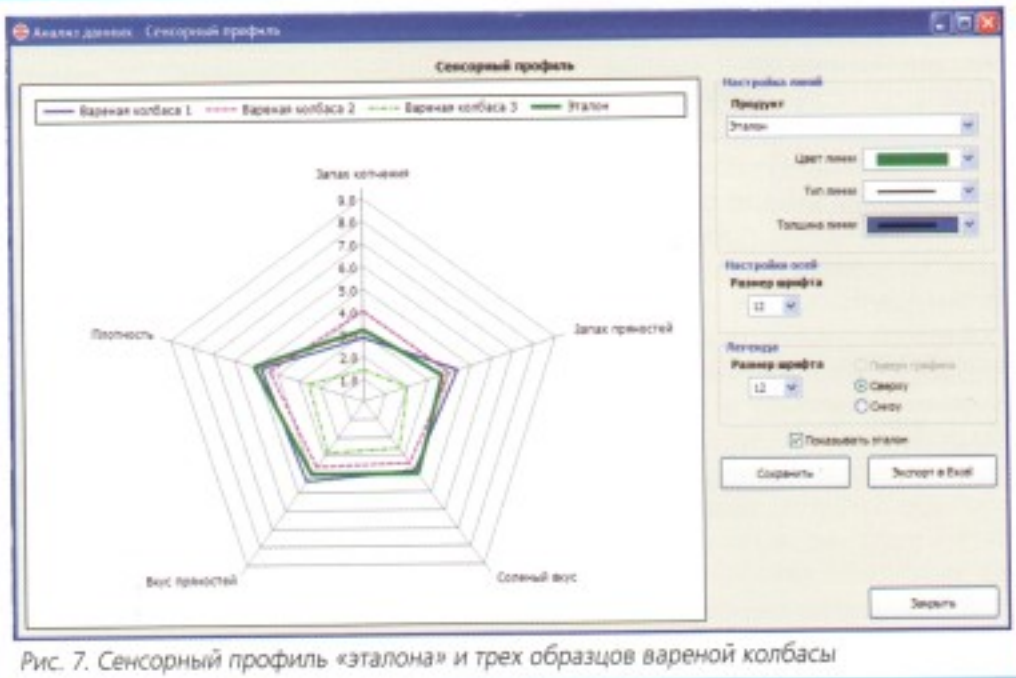

Рис. 7. Сенсорный профиль «эталона» и трех образцов вареной колбасы

представляют собой шкалы со значением 0 в центре. Количество линий равно числу оцененных дескрипторов (характеристик) продукта. На линиях откладываются отрезки, соответствующие средним арифметическим значениям оценок интенсивности.

С помощью выполнения аналогичных процедур программа дает возможность сравнить профиль продукта, вырабатываемого на предприятии с профилями продуктов конкурентов.

Для сравнения профиля продукта и «эталона», предварительно вырабатывается эталонный продукт, оцениваются его характеристики, и формируется сенсорный профиль. Для проведения сравнения загружаются данные эталонного профиля, и добавляется на график профиль аналогичного образца другой партии (рис. 7).

Разработанная программа также позволяет выявлять изменения сенсорных характеристик в продукте при замене ингредиентов или пряностей в рецептуре, использования новых видов упаковки и др.

Компьютерная программа по сбору и статистической обработке данных, основанная на принципах многофакторного дисперсионного анализа, позволяет в режиме on-line осуществлять обработку результатов дегустации. Данные, получаемые в ходе автоматической обработки, являются максимально объективными и достоверными.

ЛИТЕРАТУРА

1. *Родина, Т.Г.* Сенсорный анализ продовольственных товаров/Т.Г. Родина. – М.: Издательский центр «Академия», 2004. – 208 с.

2. *Лисицын, А.Б.* Комплексный подход к органолептической оценке как инструмент повышения качества продукции/А.Б. Лисицын, Т.Г. Кузнецова, А.А. Лазарев // Все о мясе. – 2017. – № 2. – С. 4–7

3. *Stone, H.* Sensory Evaluation Practices (3rd ed.)/H. Stone, J.L. Sidel. – San Diego: Elsevier Academic Press, 2004. – 408 p.

4. *Кантаре, В.М.* Сенсорный анализ продуктов питания/В.М. Кантаре, В.А. Матисон, М.А. Фоменко. – М.: Типография РАСХН, 2003. – 400 с.

5. *Govaert, G.* Data analysis/G. Govaert. – London: John Wiley & Sons, Inc., 2009. – 327 p.

6. *Næs, T.* Multivariate Analysis of Data in Sensory Science (1st ed.)/T. Næs, E. Risvik. – Amsterdam: Elsevier Science B.V., 1996. – 347 p.

7. *Ståhle, L.* Multivariate analysis of variance (MANOVA)/L. Ståhle, S. Wold // Chemometrics and Intelligent Laboratory Systems. – 1990. – № 9. – P. 127–141.

8. *Lee, S.M.* Development of sample preparation, presentation procedure and sensory descriptive analysis of green tea/S.M. Lee, S.-J. Chung, O.-H. Lee, H.-S. Lee, Y.-K. Kim, K.-O. Kim // Journal of Sensory Studies. – 2008. – № 23. – P. 450–467.

9. *Anderson, T.W.* An Introduction to Multivariate Statistical Analysis (4rd ed.)/T.W. Anderson. – Chichester: Wiley-Interscience, 2003. – 752 p.

10. *Heymann, H.* Comparison of canonical variate and principal component analyses/H. Heymann, A.C. Noble // Journal of Food Science. – 1989. – № 54. – P. 1355–1358.

11. *Delarue, J.* Sensory mapping using flash profile. Comparison with a conventional descriptive method for the evaluation of the flavor of fruit dairy products/J. Delarue, J.-M. Sieffermann // Food Quality and Preference. – 2004. – № 15. – P. 383–392.

12. *Luan, F.* Classification of the fragrance properties of chemical compounds based on support vector machine and linear discriminant analysis/F. Luan, H.T. Liu, Y.Y. Wen, X.Y. Zhang // Flavour and Fragrance Journal. – 2008. – № 23. – P. 232–238.

13. *Dijksterhuis, G.B.* Interpreting generalized Procrustes analysis «analysis of variance» tables/G.B. Dijksterhuis, P. Punter // Food Quality and Preference. – 1990. – № 2. – P. 255–265.

14. *Popper, R.* Three multivariate approaches to relating consumer to descriptive data. In: A.M. Muñoz (ed.). Relating Consumer, Descriptive and Laboratory Data to Better Understand Consumer Responses. ASTM Publication Code Number 28–030097–36./R. Popper, H. Heymann, F. Rossi. – West Conshohocken, PA: ASTM, 1997. – P. 39–61.

15. *Hatcher, L.* A Step-By-Step Approach to Using the SAS System for Univariate and Multivariate Statistics/L. Hatcher, P.J. Stepanski. – Cary, NC: SAS Institute, 1994. – 568 p.

16. ГОСТ Р ИСО 8586-2-2008 Органолептический анализ. Общее руководство по отбору, обучению испытателей и контролю за их деятельностью. Ч. 2. Эксперты по сенсорной оценке. – М.: Стандартинформ, 2009. – 12 с.

17. ISO 11132:2012 Sensory analysis. Methodology. Guidelines for monitoring the performance of a quantitative sensory panel. – Switzerland: International Organization for Standardization, 2006. – 28 p.

18. *Дубров, А.М.* Многомерные статистические методы/А.М. Дубров, В.С. Мхитарян, Л.И. Трошин. – М.: Финансы и статистика, 2000. – 352 с.

19. *Хруцкий, В.Е.* Современный маркетинг: настольная книга по исследованию рынка/В.Е. Хруцкий, И.В. Корнеева. – М.: Финансы и статистика, 2000. – 528 с.

20. Программа по сбору и статистической обработке сенсорных данных/М.А. Никитин [и др.] // Свидетельство о регистрации программы для ЭВМ. – М.: Роспатент. – 2017 – № 2017663406 от 01.12.2017 г.

REFERENCES

1. Rodina, T.G. Sensornyj analiz prodovol'stvennyh tovarov/T.G. Rodina. – M.: Izdatel'skij centr «Akademija», 2004. – 208 s.

2. Lisicyn, A.B. Kompleksnyj podhod k organolepticheskoj ocenke kak instrument povyshenija kachestva produkcii/A.B. Lisicyn, T.G. Kuznecova, A.A. Lazarev // Vse o mjase. – 2017. – № 2. – S. 4–7

3. Stone, H. Sensory Evaluation Practices (3rd ed.)/H. Stone, J.L. Sidel. – San Diego: Elsevier Academic Press, 2004. – 408 p.

4. Kantare, V.M. Sensornyj analiz produktov pitanija/V.M. Kantare, V.A. Matison, M.A. Fomenko. – M.: Tipografija RASHN, 2003. – 400 s.