

Хемометрические методы в контроле подлинности продуктов питания и пищевого сырья

Я.Н. Краснянчин, А.В. Пантелеймонов, Ю.В. Холин

Харьковский национальный университет имени В.Н. Каразина, пл. Свободы 4, Харьков, 61077 Украина
e-mail: kholin@univer.kharkov.ua

Поступила: 24 мая 2010 г./ Принята к публикации: 27 сентября 2010 г.

Представлен обзор основных хемометрических методов, используемых для решения задач классификации, идентификации и обнаружения при контроле подлинности продуктов питания и пищевого сырья. Наибольшее распространение для обработки массивов экспериментальных данных при идентификации продуктов питания получили следующие алгоритмы: метод главных компонент, формальное независимое моделирование аналогий классов, дискриминантный анализ, нечеткий линейный дискриминантный анализ, деревья классификации и регрессии, метод опорных векторов, проекция на латентные структуры, различные искусственные нейронные сети. Приведены примеры использования рассмотренных хемометрических алгоритмов для установления соответствия продуктов питания их заявленным наименованиям.

IA.N. KRASNIANCHYN, A.V. PANTELEIMONOV, YU.V. KHOLIN. CHEMOMETRIC METHODS IN CONTROL OF AUTHENTICITY OF FOOD PRODUCTS AND FOOD RAW MATERIALS – The authors have considered the main chemometric algorithms used to solve problems of classification, detection and identification connected with the control of authenticity of food products and food raw materials. Algorithms of principal component analysis, soft independent modeling of class analogy, discriminant analysis, fuzzy linear discriminant analysis, classification and regression trees, support vector machines, projection on latent structures and different artificial neural networks were considered. The application of chemometric algorithms for determination of such characteristics of food products and food raw materials as geographical origin, grade, content of various components and impurities was considered.

Ключевые слова: качественный химический анализ, хемометрия, продукты питания, подлинность, анализ многомерных данных.

Keywords: qualitative chemical analysis, chemometrics, food products, authenticity, multivariate data analysis.

ВВЕДЕНИЕ

В 2005 году авторитетный журнал "Trends in Analytical Chemistry" посвятил один из своих номеров [1] проблематике современного качественного химического анализа. Редакторы выпуска М. Валкарсел и С. Карденас указывали [1, Р. 467], что в последние десятилетия качественный анализ считали чуть ли не умирающим разделом аналитической химии, по традиции связанным лишь с обнаружением катионов и анионов в растворах; ситуация кардинально изменилась в последние годы, когда стало ясно, что значимые для практики качественные заключения могут быть основаны не только на сенсорном восприятии признаков химических реакций, но и сделаны на основе обработки массивов количественных данных, получаемых инструментальными методами.

К основным задачам качественного химического анализа относят идентификацию аналитов и их обнаружение [2]. Под "идентификацией" понимают заключение о тождественности анализируемого и известного объектов или о принадлежности аналита некоторому классу объектов на основе соот-

ветствия их свойств [3]¹. "Обнаружение" – вывод о присутствии определенного аналита в исследуемой пробе в концентрации, превышающей заранее установленное пороговое значение. При решении задач обнаружения отвечают на вопросы "Загрязнено ли попавшее в продажу подсолнечное масло *нефтью*?", "Принимал ли спортсмен *определенный запрещенный препарат*?", "Превышает ли в минеральной воде содержание *определенного токсиканта* ПДК?" и т.п. И для идентификации, и для обнаружения используют аналитические процедуры двух типов: а) непосредственно приводящие к тому или иному заключению (методики с

¹ Характерны такие, например, определения: "Идентификация – это установление вида и состояния молекул, ионов, радикалов, атомов и др. частиц на основе сопоставления экспериментальных данных с соответствующими справочными данными для известных частиц" [4]; "Chemical identification is considered to be assigning an analyte (analytical signal) to one of the set of known individual chemical compounds or to a group/class of compounds based on matching their properties" [2].

бинарным откликом) и б) формирующие заключение на основе обработки массивов первичных количественных данных, полученных инструментальными методами [5].

В настоящее время все более отчетливо проявляется тенденция к расширительной трактовке качественного анализа как процедуры классификации объектов по их признакам ("The classification of objects against specified criteria to meet an agreed requirement" [6]²). В таком подходе естественным образом объединяются задачи идентификации и обнаружения, а результаты анализа рассматриваются как рекомендация для принятия потребителем (заказчиком анализа) управленческих решений. Процедуры качественного анализа могут использовать классификацию как с обучением, так и без обучения, и, поскольку речь обычно идет об анализе массивов многомерных данных, требуют применения хемометрических процедур.

Отдельная проблема – метрология качественного химического анализа, принципы которой существенно отличаются от метрологии количественного анализа. Действительно, центральная характеристика результата количественного анализа – его неопределенность – в качественном анализе аналогов не имеет.

Вместо этого результаты качественного анализа характеризуют таким показателем, как ненадежность (unreliability). Ненадежность – это количественная оценка вероятности ложного логического заключения (ложной классификации). Примерами могут служить ложная идентификация сложного эфира как амида, вывод об отсутствии аналита, который на самом деле наличествует в пробе, неправомерное отнесение образца минеральной воды к группе вод с высокой минерализацией и т.п. В метрологии количественного анализа оценивание неопределенности проработано достаточно глубоко (см., например, [7]), тогда как подходы к оцениванию ненадежности результатов качественного анализа остаются дискуссионными.

Ренессанс качественного анализа обусловлен быстрым ростом практических запросов на массовый анализ проб сложного состава в новых предметных областях (клинический анализ, промышленная гигиена, контроль психотропных и наркотических препаратов, оценка токсичности объектов окружающей среды, сигнальный контроль загрязнителей, токсикантов, диверсионных

ядов и др.). Одной из таких областей является идентификация и обеспечение безопасности пищевых продуктов, напитков и пищевого сырья.

Под идентификацией пищевых продуктов понимают установление соответствия продукта его заявленному наименованию (виду, классу, категории, сорту, географическому происхождению и т.п. [8–19]).

Идентификация пищевых продуктов, согласно национальному стандарту Украины 4161-2003 "Системы управления безопасностью пищевых продуктов. Требования", и национальному стандарту 22000:2007 "Системы управления безопасностью пищевых продуктов. Требования ко всем организациям пищевой цепи", гармонизированному со стандартом ISO, является частью системы обеспечения качества пищевых продуктов.

Продукты, не соответствующие заявленному наименованию, относят к разряду фальсифицированных. "Фальсификацией" (нарушением подлинности) считают подделывание чистого или подлинного товара или замену дорогостоящего компонента компонентом с низкой стоимостью для получения незаконной дополнительной выгоды, что ведет к снижению качества продукта.

К нарушению подлинности продуктов могут приводить следующие действия [20, 21]:

1. полная или частичная замена компонентов;
2. полное или частичное упущение или извлечение ценных элементов;
3. дополнение необъявленного вещества или материала с целью увеличения веса, снижения качества или улучшения внешней привлекательности продукта;
4. несоблюдение норм технологического процесса;
5. несоблюдение законных требований (стандарт продукта, географическое происхождение продукта, максимальное или минимальное содержание воды и др.).

Следовательно, в отношении продуктов питания термины "подлинность" или "аутентичность" обозначают их неподдельность, натуральность, соответствие указанным в сертификатах сортовому, видовому и географическому происхождению, а также технологии их переработки, отсутствие нерегламентированных примесей и добавок [22].

Сопоставление показателей исследуемого продукта и аутентичных образцов (эталонов), их описаний, опубликованных в соответствующих документах, а также информации, содержащейся в сопроводительных документах и потребительских этикетках, с применением аналитических и органолептических методов является быстро раз-

² Для того чтобы избежать чересчур расширительной трактовки (скажем, исключить из процедур качественного химического анализа подсчет результатов заполнения бюллетеней для голосования) уместно ограничить перечень признаков химическими или физико-химическими характеристиками.

вивающейся и актуальной областью качественно-го химического анализа [8, 23–34].

О растущем внимании к подлинности пищевых продуктов как к научной проблеме свидетельствует и увеличение числа соответствующих источников, реферируемых системой SCOPUS (рис. 1, табл. 1).

Если вывод о тождественности анализируемого продукта и аутентичного образца или об отношении продукта к определенному классу предстоит сделать на основе обработки многомерных массивов экспериментальных данных, неизбежным становится применение хемометрических методов, среди которых особое значение имеют алгоритмы классификации, распознавания образов, дискриминантного анализа и искусственных нейронных сетей [35–41] (табл. 2).

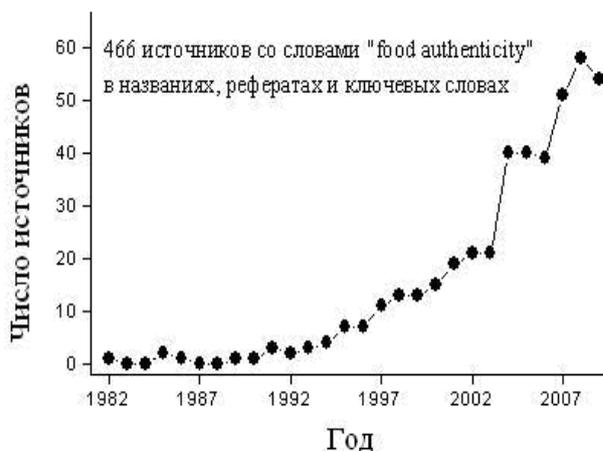


Рис. 1. Данные об источниках в системе SCOPUS, относящихся к аутентичности пищевых продуктов и сырья.

Таблица 1. Число источников в системе SCOPUS, содержащих в названиях, рефератах и ключевых словах название пищевого продукта и термин "authenticity"

Продукт	Всего источников	Год											
		2000	2001	2002	2003	2004	2005	2006	2007	2008	2009	2010*	
Растительное масло	188	7	7	22	11	8	15	17	17	23	19	16	
Фрукты	141	6	9	6	7	9	11	11	10	11	14	6	
Вино	94	4	4	4	2	6	6	11	10	15	9	3	
Маслины	70	3	2	9	3	3	3	4	10	9	5	7	
Мед	63	2	4	3	3	5	2	9	5	9	5	4	
Молоко	66	1	3	3	6	7	8	4	6	5	8	5	
Мясо	63	2	2	3	5	6	3	3	2	6	8	7	
Рыба	54	1	3	5	2	2	6	4	5	7	6	3	
Сыр	44	0	1	3	6	5	5	4	3	3	7	3	

* по состоянию на 26 июня 2010 г.

Таблица 2. Число источников в системе SCOPUS, содержащих в названиях, рефератах и ключевых словах термин "food products" и название хемометрического метода

Метод	Год											
	2000	2001	2002	2003	2004	2005	2006	2007	2008	2009	2010*	
PCA	18	19	30	44	69	56	56	75	97	54	34	
SIMCA	1	0	1	0	1	1	1	4	0	1	1	
DA	9	6	9	14	14	12	18	27	40	14	10	
CART	0	0	0	1	1	1	1	0	3	1	0	
SVM	0	0	0	1	1	4	4	9	15	2	3	
PLS	0	0	0	1	0	0	0	0	4	0	0	
ANN	8	6	9	13	16	15	17	31	27	14	10	
Всего						954						

* по состоянию на 17 июня 2010 г.

Примечание:

- PCA – метод главных компонент;
- SIMCA – формальное независимое моделирование аналогий классов;
- DA – дискриминантный анализ;
- CART – деревья регрессии и классификации;
- SVM – метод опорных векторов;
- PLS – проекция на латентные структуры;
- ANN – искусственные нейронные сети.

В научной периодике публикации по исследованию подлинности пищевых продуктов (в частности, с использованием хеометрических методов) сосредоточены, главным образом, в журналах хемико-аналитического профиля и изданиях, посвященным химии пищевых продуктов, сырья и про-

цессов (табл. 3), причем весьма значительная доля работ выполнена авторами из стран, для которых производство и переработка сельскохозяйственного сырья, пищевая индустрия являются важными отраслями экономики – Испании, Италии, Франции, Греции.

Таблица 3. Основные периодические издания, публикующие работы по применению хеометрических методов при анализе подлинности пищевых продуктов и сырья (по данным SCOPUS за 2000–2010 гг. при запросе "food & chemometrics")

Химические издания		Нехимические издания	
Название	Число источников*	Название	Число источников*
Anal. Chim. Acta	35	J. of Agricultural and Food Chem.	58
Chemometr. and Intell. Lab. Syst.	17	Food Chem.	31
J. of Chromatography A	16	J. of Sci. of Food and Agriculture	9
J. of Near Infrared Spectroscopy	10	J. of Dairy Sci.	8
Anal. and Bioanal. Chem.	10	Critical Rev. in Food Sci. and Nutrition	7
Analyst	5	J. of Food Sci.	6
J. of Chemometrics	7	Int. J. of Food Sci. and Technol.	6
Appl. Spectroscopy	7	Trends in Food Sci. and Technol.	4
Analyt. Chem.	5		
Trends Anal. Chem.	4		
Talanta	4		
Critical Rev. in Anal. Chem.	3		

* по состоянию на 2 июля 2010 г.

Несмотря на большое значение этих отраслей и для Украины и России, а также многочисленные факты фальсификации пищевых продуктов в наших странах, российских работ по изучению качества и проверке подлинности продуктов и напитков довольно мало (см., например, [9, 42–48]), а украинские работы по этой тематике в международной периодике отсутствуют. Можно полагать, что это связано со слабостью приборной базы, необходимой для получения больших массивов данных о характеристиках анализируемых объектов, а также с отставанием от международного уровня исследований в области хеометрии.

Поскольку при контроле подлинности пищевых продуктов и сырья выполнение измерений инструментальными методами и обработка их результатов с помощью хеометрических процедур неразделимы, представляется уместным представить обзор хеометрических методов, получивших наибольшее распространение в данной области.

ЭКСПЕРИМЕНТАЛЬНЫЕ МЕТОДЫ

В анализе пищевых продуктов и сырья используют большой набор методов, обеспечивающих получение массивов экспериментальных данных, дальнейшую обработку которых проводят с помощью хеометрических процедур (табл. 4).

Недавно опубликованы подробные обзоры экспериментальных методов получения многомерных массивов данных, используемых для хеометрического исследования качества и проверки подлинности пищевых продуктов (на примере молочных продуктов и томатов) [21, 49]. Поэтому в данной работе ограничимся лишь кратким обзором важнейших инструментальных методов.

Экспрессность и чувствительность хроматографического анализа, возможность сочетания хроматографии с другими физико-химическими методами [50–55] сделали его наиболее распространенной процедурой анализа продуктов питания и напитков.

Различные виды хроматографии используются на всех стадиях производственного процесса пищевых продуктов и в контроле их качества: высокоэффективная жидкостная хроматография (дополняемая капиллярным электрофорезом), ионная, газовая, тонкослойная, мицеллярная электрокинетическая хроматография.

Хроматографические методы успешно применяли для контроля качества безалкогольных и алкогольных напитков, молочных продуктов, меда, искусственных пищевых добавок, выявления загрязнителей антропогенного и природного характера [42, 44–46, 56–66].

Таблица 4. Число источников в системе SCOPUS, содержащих в названиях, рефератах и ключевых словах термин "food products" и название инструментального метода

Метод	Год											
	2000	2001	2002	2003	2004	2005	2006	2007	2008	2009	2010*	
HPLC	136	138	183	185	283	241	275	258	344	243	107	
IC	2	2	0	2	5	1	3	5	7	6	1	
GC	112	129	164	178	235	231	197	232	246	175	73	
TLC	18	19	28	28	40	38	33	25	31	21	11	
MEC	1	6	2	5	3	10	5	8	7	2	3	
GLC	8	5	6	7	10	8	5	9	9	5	3	
EN	8	9	10	9	14	12	14	16	11	14	5	
ET	1	0	3	0	2	4	4	1	8	2	2	
VS	0	0	3	1	3	1	2	1	1	0	1	
NIRS	7	10	10	19	33	34	30	34	48	33	15	
FTIS	11	12	11	15	16	25	21	31	33	30	13	
RS	2	0	5	2	9	17	8	10	14	10	8	
NMR	24	25	44	56	64	53	63	61	64	60	27	
MS	114	127	201	234	330	323	316	367	380	341	147	
IRMS	2	1	3	7	5	11	11	10	6	4	5	
ICP-MS	5	5	6	14	17	15	10	15	28	20	3	
AAS	8	12	17	15	15	12	14	10	16	12	5	
LS	0	1	0	1	2	0	0	1	5	1	1	
Всего	9328											

* по состоянию на 17 июня 2010 г.

Примечание:

- HPLC – высокоэффективная жидкостная хроматография;
- IC – ионная хроматография;
- GC – газовая хроматография;
- TLC – тонкослойная хроматография;
- MEC – мицеллярная электрокинетическая хроматография;
- GLC – газо-жидкостная хроматография;
- EN – "электронный нос";
- ET – "электронный язык";
- VS – видимая спектроскопия;
- NIRS – инфракрасная спектроскопия в ближнем ИК-диапазоне;
- FTIS – инфракрасная спектроскопия с Фурье-преобразованием;
- RS – Рамановская спектроскопия;
- NMR – ядерный магнитный резонанс;
- MS – масс-спектрометрия;
- IRMS – изотопная масс-спектрометрия;
- ICP-MS – масс-спектрометрия с индуктивно связанной плазмой;
- AAS – атомно-абсорбционная спектрофотометрия;
- LS – люминесцентная спектроскопия.

Различные спектроскопические методы (видимая спектроскопия [67], инфракрасная спектроскопия в ближнем ИК-диапазоне [67, 68], инфракрасная спектроскопия с Фурье-преобразованием [69], Рамановская спектроскопия [70], ядерный магнитный резонанс [71, 72], масс-спектрометрия [73] (в том числе изотопная масс-спектрометрия [52, 74–77] и масс-спектрометрия с индуктивно связанной плазмой [37]), атомно-абсорбционная спектрофотометрия [78–80], люминесцентная спектроскопия [81]) также интенсивно используются при исследовании подлинности продуктов пи-

тания. Относительная простота подготовки образцов и выполнения измерительных процедур, возможность довольно просто получать обширные массивы данных о свойствах анализируемых объектов способствовали широкому внедрению спектроскопических методов; при проверке идентичности и классификации образцов спектральные данные часто используют как "отпечатки пальцев" объектов ("fingerprints") [15, 82, 83].

Создание таких сенсорных систем, как "электронный нос" (для анализа газообразных образцов) и "электронный язык" (для анализа жидких

образцов) было вызвано стремлением реализовать аналитические процедуры, моделирующие обонятельные и дегустационные способности человека. Системы "электронного языка" и "электронного носа" состоят из набора сенсоров различной селективности. С их использованием получают многомерные массивы характеристик анализируемых образцов, а затем с помощью хемометрических методов получают информацию об исследуемом образце, соотносимую с человеческим восприятием, например, вкуса образца (соленость, кислотность, горечь), сведениями о его географическом происхождении, составе, интенсивности аромата и степени свежести.

Распространены химические сенсоры различных типов: вольтамперометрические, потенциометрические, спектрофотометрические, флуоресцентные, кондуктометрические, амперометрические и др.

Растущая популярность систем "электронного носа" и "электронного языка" связана с возможностями их миниатюризации, автоматизации работы, простотой применения и экспрессностью. В обработке массивов данных, получаемых на основе использования наборов сенсоров различной селективности, широкое применение нашли современные хемометрические методы, в частности, методы распознавания образов и искусственных нейронных сетей.

С учетом достижений теории нейронных вычислений идентификацию объектов анализа с помощью "электронного языка" и "электронного носа" можно рассматривать как ветвь развития искусственного интеллекта и/или область применения "электронного мозга" [84–93].

ХЕМОМЕТРИЧЕСКИЕ МЕТОДЫ

Развитие хемометрии быстро пополняет арсенал доступных химикам-аналитикам алгоритмов обработки экспериментальных данных. Многие подходы реализованы как в специализированных пакетах программ (Unscrambler, SIMCA, Eigenvector Research и др.), так и в статистических пакетах общего назначения (Statistical Package for the Social Sciences, Statistica, Matlab).

Метод главных компонент (Principal component analysis, PCA)

Для предварительной обработки и извлечения информации из многомерных массивов экспериментальных данных используются методы сжатия данных. Сжатие позволяет представить результаты измерений в компактном виде, удобном для визуализации и интерпретации. Наиболее распространенной хемометрической технологией понижения размерности массива данных с минималь-

ной потерей информации является метод главных компонент.

Метод главных компонент используется в анализе любых сложных химических данных, для которых характерна мультиколлинеарность, т.е. присутствие внутренних, скрытых связей между переменными. Основными предпосылками для того, чтобы эффективно снизить размерность массива данных, являются "сильная" (вплоть до линейной) связь между исходными переменными, в результате которой информация, содержащаяся в данных, дублируется; слабая информативность некоторых показателей, что позволяет исключить их из набора данных; возможность объединения нескольких показателей в один [94, 95].

Алгоритм осуществляется итерационно и его цель – расчет новых переменных, ортогональных и некоррелированных главных компонент. Метод главных компонент применяется к данным, записанным в виде матрицы X размерностью $I \times J$, строки которой соответствуют анализируемым образцам ($i = 1, \dots, I$), а столбцы – исходным переменным $x_j (j = 1, \dots, J)$, характеризующим эти образцы. Главные компоненты $t_a (a = 1, \dots, A)$ – линейная комбинация исходных переменных:

$$t_a = p_{a1}x_1 + \dots + p_{aJ}x_J. \quad (1)$$

С помощью этих новых переменных исходный массив данных X разлагается в произведение двух матриц T и P :

$$X = TP^t + E = \sum_{a=1}^A t_a p_a^t + E.$$

В этом уравнении T (размерность $I \times A$) – матрица счетов (scores), P (размерность $J \times A$) – матрица нагрузок (loadings), E (размерность $I \times J$) – матрица остатков. Метод главных компонент можно понимать как проецирование данных на подпространство (главных компонент) с размерностью меньшей, чем исходное пространство. Строки t_1, \dots, t_I матрицы T – это координаты образцов в новой системе координат, столбцы t_1, \dots, t_A матрицы T представляют проекции всех образцов на одну новую координатную ось. Каждая строка матрицы P состоит из коэффициентов, связывающих переменные t и x (1), каждый столбец P – это проекция соответствующей переменной x_j на новую систему координат [96].

В результате применения метода главных компонент осуществляется переход к новому ортогональному базису, оси которого ориентированы по направлениям максимальной дисперсии набора

входных данных. Такое преобразование позволяет сжимать информацию путем отбрасывания координат, соответствующих направлениям с минимальной дисперсией. Каждая главная компонента имеет дисперсию, максимально возможную из всех комбинаций исходных переменных, при условии ортогональности предыдущей главной компоненте. Таким образом, задача сводится к поиску таких линейных комбинаций исходных показателей, которые объясняли бы максимально возможную долю изменчивости (сумму дисперсий) исходных показателей. В большинстве случаев многопараметрический массив экспериментальной информации, содержащей десятки переменных, можно представить в виде массива, содержащего две-три переменные, т.к. первые две или три главные компоненты отражают около 80% исходных данных. Возникающие при этом остатки (матрица E) рассматриваются как шум, не содержащий значимой химической информации [97–99].

Как правило, методу главных компонент предшествует центрирование (вычитание из каждого столбца x_j среднего по столбцу значения m_j) или нормирование (деление каждого столбца x_j на собственное стандартное отклонение s_j) исходной матрицы X , т.к. величины дисперсий показателей существенно зависят от масштаба единиц измерения [96]:

$$m_j = (x_{1j} + \dots + x_{lj}) / I ;$$

$$s_j = \sqrt{\sum_{i=1}^I (x_{ij} - m_j)^2 / I} .$$

Метод главных компонент широко применяется в исследовании подлинности пищевых продуктов. С его помощью стараются выявить взаимосвязи данных, составляющих большие массивы результатов первичных измерений. Получение таких массивов экспериментальных данных характерно для использования технологии "электронного языка" [71, 100, 101], методов хроматографии [13, 102], инфракрасной спектроскопии в ближнем диапазоне [103], ядерного магнитного резонанса [104] и др.

В качестве примера рассмотрим применение метода главных компонент при контроле качества цветочных видов меда.

В работе [102] исследовано содержание сахаров в 280 образцах французского меда семи цветочных разновидностей (акация, каштан, рапс, лаванда, ель, липа, подсолнух). С помощью газовой

хроматографии, с пламенным ионизационным детектором, оценивали содержание 17-ти ди- и трисахаридов. А с помощью жидкостной хроматографии, с импульсным амперометрическим детектором, определяли содержание глюкозы и фруктозы (табл. 5).

Подлинность цветочного происхождения 280 образцов меда подтвердили на основании результатов пыльцевого, органолептического и лабораторного анализов. Авторы поставили задачу определить такие характеристики образцов меда, которые обеспечивали бы их надежную идентификацию.

Для достижения поставленной цели применили метод главных компонент. Для обработки методом главных компонент из полученного массива первичных экспериментальных данных с помощью дисперсионного анализа (analysis of variance, ANOVA) отобрали такие характеристики, как массовые концентрации глюкозы, раффинозы, трехалозы, фруктозы, и отношение концентраций эрлоза/мальтулоза, мальтоза/трехалоза и фруктоза/глюкоза. Набор из этих характеристик обеспечивал идентификацию 72% образцов меда.

Затем к массиву переменных, отобранных с помощью дисперсионного анализа, применили метод главных компонент.

В результате работы выделили две главные компоненты, являющиеся линейной комбинаций вышеназванных характеристик. Таким образом, массив данных, включавший 20 характеристик образцов меда семи цветочных разновидностей, был сжат до массива размером 2×7 .

Метод главных компонент позволил определить границы областей каждой цветочной разновидности меда, например, область образцов елового меда определилась четко, но при этом области других цветочных разновидностей меда перекрылись: акациевые и каштановые, лавандовые и липовые, рапсовые и подсолнуховые.

Фактически, несмотря на отсутствие полной дифференциации, такой результат дал возможность контролировать качество меда в соответствии с его цветочной разновидностью.

Предложенный алгоритм применили для оценки подлинности 47 коммерческих образцов меда: 34 французских (8 акациевых, 8 каштановых, 7 лавандовых, 6 еловых, 5 липовых), 7 венгерских и 3 китайских (акациевых), 1 испанского и 1 марокканского (лавандовых), 1 турецкого (елового).

С его помощью обнаружили, что 5 акациевых и 4 еловых коммерческих образца меда находятся за пределами их предполагаемой области, а, значит, не соответствуют своему заявленному виду.

Таблица 5. Среднее содержание и стандартное отклонение (M, SD соответственно) содержания сахаров в 280 образцах меда семи цветочных разновидностей

Число образцов	Акация		Каштан		Ряпс		Лаванда		Ель		Липа		Подсолнух	
	50		38		28		53		37		38		36	
	M ^a	SD	M	SD	M	SD	M	SD	M	SD	M	SD	M	SD
Фруктоза	438,9	12,9	407,2	22,8	379,2	15,2	384,6	13,0	314,9	19,9	381,6	26,3	399,8	22,6
Глюкоза	263,0	13,0	265,2	20,7	396,4	22,9	323,3	13,1	241,7	25,1	326,7	19,0	379,0	18,5
Ф/Г ^b	1,7	0,1	1,5	0,1	1,0	0,1	1,2	0,1	1,3	0,1	1,2	0,1	1,1	0,1
Сахароза	2,0	2,0	0,2	0,2	0,1	0,1	4,7	2,8	0,4	0,3	0,5	0,6	0,1	0
Мальтоза	2,6	0,7	1,5	0,5	0,7	0,3	2,6	0,6	1,7	0,4	1,9	0,4	1,1	0,4
Мальтулоза	1,7	0,4	2,6	0,9	0,9	0,4	1,0	0,3	1,9	0,6	2,3	1	0,5	0,2
Тураноза	2,9	0,6	2,8	0,8	1,3	0,6	1,8	0,5	2,5	0,4	3,2	1,1	0,9	0,3
Трехалоза	1,5	0,4	2,0	0,7	0,6	0,3	0,8	0,2	4,5	1,1	1,9	0,6	0,6	0,2
Палатиноза	0,3	0,1	0,5	0,4	0,2	0,1	0,2	0,1	0,5	0,4	0,4	0,2	0,1	0,0
Ламинарибиоза	1,2	0,3	1,5	0,5	0,5	0,3	0,7	0,3	1,3	0,3	1,3	0,4	0,5	0,2
Мелибиоза	0,1	0,1	0,3	0,3	НО ^c		0,1	0,0	0,3	0,1	0,3	0,1	НО	
Изомальтоза	0,9	0,3	1,8	0,8	0,4	0,2	0,5	0,2	1,8	0,6	1,7	0,8	0,3	0,1
Гентиобиоза	0,0	0,0	0,2	0,5	НО		НО		0,1	0,0	0,2	0,2	НО	
Раффиноза	0,0	0,1	0,0	0,1	НО		0,1	0,0	2,1	0,6	НО		НО	
Неокестоза	0,2	0,1	0,2	0,3	0,0	0,1	0,2	0,0	0,4	0,5	0,4	0,2	0,1	0,0
1-Кестоза	0,1	0,0	0,1	0,1	НО		0,1	0,0	НО		0,3	0,1	НО	
Эрлоза	1,9	1,2	0,2	0,2	НО		1,4	0,6	2,1	1,3	1,0	0,6	НО	
Мелезитоза	0,1	0,1	0,2	0,4	НО		0,1	5,7	4,2	0,2	0,1	НО		
Мальтотриоза	0,4	0,2	0,2	0,1	НО		0,2	0,1	0,8	0,3	0,5	0,2	0,1	0,0
Рамноза	0,2	0,1	0,2	0,1	0,1	0,2	0,1	0,1	0,3	0,2	0,4	0,2	0,1	0,1

^a В процентных единицах (кроме глюкозы и фруктозы = г/кг)

^b Фруктоза/глюкоза

^c Не обнаружено

Формальное независимое моделирование аналогий классов (Soft independent modeling of class analogy, SIMCA)

Формальное независимое моделирование аналогий классов методом распознавания образов с обучением – хорошо известное и широко применяемое хемометрическое средство решения задач классификации [82, 105]. Классификация объектов с помощью SIMCA происходит в два этапа: этап обучения и этап тестирования. На первом этапе формируются отдельные модели классов. Каждый класс из обучающего множества объектов независимо моделируется методом главных компонент с различным числом главных компонент A [94, 106]. Под моделью в данном случае понимают

способ воспроизведения формы и/или других характеристик сложного исходного массива в более простой форме, включающей также и набор данных, характеризующих свойства моделируемой системы, и динамику их изменения.

Для сравнения созданных моделей рассчитывают расстояние между классами и оценивают влияние переменных на разделение по классам (модельная мощность и дискриминационная мощность переменных) [107, 108].

Значение модельной мощности переменной (M_j) показывает, насколько сильное влияние данная переменная оказывает на моделирование данного класса (1 – сильное влияние, 0 – влияния нет):

$$M_j = 1 - \frac{S_{jresid}}{S_{jraw}},$$

где S_{jraw} – стандартное отклонение исходных значений переменной, S_{jresid} – стандартное отклонение остатков, определяемых методом главных компонент для соответствующей модели данного класса. Значение дискриминационной мощности переменной (D_j) показывает способность переменной разделять классы (способность переменной моделировать класс не влечет за собой способность к разделению). Пусть мы имеем по две модели для классов X_1 и X_2 , тогда значение дискриминационной мощности интересующей нас переменной имеет вид:

$$D_j = \sqrt{\frac{S_{jresid}^2(classX_1model B) + S_{jresid}^2(classX_2model A)}{S_{jresid}^2(classX_1model A) + S_{jresid}^2(classX_2model B)}}.$$

Чем больше значение дискриминационной мощности переменной, тем больше ее способность к разделению классов.

Для вычисления расстояния между классами используют расстояние Махаланобиса; чем больше значение данного параметра, тем лучше различаются модели:

$$H = (\mu_1 - \mu_2)^T \cdot cov^{-1}(X) \cdot (\mu_1 - \mu_2),$$

где μ_1, μ_2 – векторы средних значений переменных для классов X_1, X_2 , $cov(X)$ – объединенная ковариационная матрица для классов X_1, X_2 ,

$$cov(X) = \frac{cov(X_1) + cov(X_2)}{(n_1 + n_2 - 2)},$$

где n_1, n_2 – длины векторов μ_1, μ_2 .

На этапе тестирования новые объекты (не используемые на стадии обучения) относят к классам, сформированным на первом этапе. Этап классификации новых образцов включает расчет расстояний от образца до центра класса (H_1) и от образца до класса (H_2) [96, 107–111]:

$$H_1 = \frac{1}{I} + \sum_{a=1}^A \frac{\tau_a^2}{t_a^T t_a},$$

$$H_2 = \sqrt{\frac{1}{J-A} \sum_j E_j^2},$$

где I, J – число образцов и число переменных соответственно, τ_a – проекция нового образца на

главную компоненту a , t_a – вектор, содержащий счета всех обучающих образцов в классе.

Расстояние от нового образца до класса показывает, насколько далеко образец находится от данного класса; расстояние от образца до центра класса вычисляется как *размах* и показывает, насколько проекция образца на данный класс далека от его центроида, т.е. насколько он отличается от других образцов данного класса.

Такие моделирующие свойства и возможность отнесения образца к одному или нескольким классам или к ни одному из смоделированных классов (в отличие от классических методов классификации, которые позволяют отнести образец только к одному классу) обеспечили методу формального независимого моделирования аналогий классов популярность при определении подлинности продуктов питания, в особенности в комбинации со спектроскопическими методами [110–117].

Примером успешного применения алгоритма SIMCA служит работа [82], в которой его использовали для классификации вин по их виду, происхождению винограда и технологии выдержки. Для этого получили "отпечатки пальцев" (нормализованные спектры в средней ИК-области) 60 образцов красных вин и 60 образцов белых вин. Затем рассчитали матрицы подобия между спектрами на основе индекса подобия Танимото:

$$T_{A,B} = \frac{c}{a+b-c},$$

где c – множество фрагментов (частиц) "отпечатков пальцев", общих для соединений A и B ; a – множество фрагментов "отпечатков пальцев" соединения A , отсутствующих у соединения B ; b – множество фрагментов "отпечатков пальцев" соединения B , отсутствующих у соединения A .

Ошибки классификации по типу вина (белые или красные) в результате применения SIMCA к спектрам образцов и к матрицам подобия спектров составили соответственно 20% и 10%. Отношение числа ошибочно отнесенных объектов к общему числу объектов служит простейшей статистической (эвристической) оценкой ненадежности идентификации.

При применении SIMCA к матрицам спектрального подобия для различения вин по виду винограда и способу выдержки ошибки составили 9% и 5%, соответственно.

Итогом работы стала демонстрация возможности создать библиотеку "отпечатков пальцев" различных образцов вин.

Дискриминантный анализ (Discriminant analysis, DA)

Дискриминантный анализ является разделом многомерного статистического анализа, который включает в себя методы классификации многомерных наблюдений по принципу максимального сходства при наличии обучающих признаков. Основными задачами дискриминантного анализа являются исследование межклассового различия (дискриминации) в априорно заданных классах по нескольким переменным одновременно, т.е. определение "вклада" каждой из переменных в различение классов, а также классификация объектов, не входивших в обучающую выборку. Переменные, которыми оперируют для поиска различий между классами, называют дискриминантными переменными. В дискриминантном анализе находят такую комбинацию дискриминантных переменных и дискриминантную функцию, которые бы оптимально разделяли рассматриваемые классы [106–118].

Выполнение дискриминантного анализа подразумевает наличие минимум двух классов и минимум двух объектов в каждом классе, линейную независимость дискриминантных переменных, многомерную нормальность закона распределения дискриминантных переменных для каждого класса, приблизительное равенство ковариационных матриц для всех классов.

Основными проблемами являются отбор дискриминантных переменных и выбор вида дискриминантной функций (линейная или нелинейная). Наиболее часто используется линейная форма дискриминантной функции следующего вида [119]:

$$d_{km} = \beta_0 + \beta_1 x_{1km} + \dots + \beta_p x_{pkm}; \\ m = 1, \dots, n; k = 1, \dots, g,$$

где d_{km} – значение дискриминантной функции для m -го объекта в классе k , x_{ikm} – значение дискриминантной переменной x_i для m -го объекта в классе k , β_i – дискриминантные множители.

Кратко опишем процедуру дискриминантного анализа при использовании линейной дискриминантной функции (linear discriminant analysis, LDA) [119–123].

Выбор дискриминантных переменных можно проводить методом пошагового включения или пошагового исключения переменных. Процедура последовательного включения начинается с выбора переменной, обеспечивающей наилучшее одномерное различие. Затем анализируют пары, образованные отобранной переменной и одной из оставшихся, в результате чего находят пару, дающую наилучшее различие. На каждом

следующем шаге процедуры отбирают переменные, которые в сочетании с отобранными ранее дают наилучшее различие. Процедура последовательного исключения исходит из предположения, что все переменные считаются входящими в систему, а затем на каждом шаге отбрасывается одна, обеспечивающая в сочетании с оставшимися переменными наихудшее различие.

Процедура пошагового дискриминантного анализа предполагает проверку (в начале каждого шага) всех дискриминантных переменных на соответствие двум условиям: необходимой точности вычисления (толерантности) и превышения заданного уровня различия (по F -критерию Фишера). Статистика F -ввода оценивает улучшение различия благодаря использованию данной переменной по сравнению с различием, достигнутым с помощью уже отобранных переменных. Статистика F -исключения определяет значимость ухудшения различия после удаления переменной из списка уже отобранных переменных. Как известно, критерий Фишера применяется для проверки равенства дисперсий двух выборок. Если вычисленное значение F -критерия больше критического для определенного уровня значимости и соответствующего числа степеней свободы для числителя и знаменателя, то дисперсии считаются различными:

$$F = \frac{\sigma_1^2}{\sigma_2^2},$$

где σ_1^2 – большая, σ_2^2 – меньшая дисперсии.

Значение толерантности переменной определяется как $(1 - R_i^2)$, где R_i^2 – квадрат коэффициента множественной корреляции i -той переменной со всеми остальными переменными (предположим, что мы рассчитываем толерантность для переменной y , причем переменные x, z уже отобраны для различения классов) [124, 125]:

$$L = 1 - \frac{r_{yx}^2 + r_{yz}^2 - 2r_{yx}r_{yz}r_{xz}}{1 - r_{xz}^2},$$

где r_{yx}, r_{yz}, r_{xz} – простые коэффициенты линейной парной корреляции;

$$r_{yx} = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^n (x_i - \bar{x})^2 \sum_{i=1}^n (y_i - \bar{y})^2}}.$$

Значение толерантности служит мерой избыточности переменной. Например, если переменная

ная, предназначенная для включения в дискриминантную функцию, имеет значение толерантности $L = 0,01$, то она может рассматриваться как на 99% избыточная по сравнению с уже включенными переменными.

Коэффициенты β_i первой дискриминантной функции выбираются таким образом, чтобы центры различных классов как можно больше отличались друг от друга. Коэффициенты второй функции выбираются так же, но при этом значения второй функции должны быть некоррелированными со значениями первой. Аналогично определяются и другие функции.

Для получения коэффициентов β_i дискриминантной функции нужен статистический критерий различения классов. Классификация переменных будет осуществляться тем лучше, чем меньше рассеяние точек относительно центра внутри группы и чем больше расстояние между центрами групп. Таким образом, метод поиска наилучшей дискриминантной функции заключается в нахождении такой дискриминантной функции, которая бы максимизировала отношение межгрупповой вариации (матрица B) к внутригрупповой (матрица W). Элементы этих матриц определяются выражениями:

$$w_{ij} = \sum_{k=1}^g \sum_{m=1}^{n_k} (x_{ikm} - \bar{x}_{ik})(x_{jkm} - \bar{x}_{jk}), \quad (2)$$

$$b_{ij} = \sum_{k=1}^g n_k (\bar{x}_{ik} - \bar{x}_i)(\bar{x}_{jk} - \bar{x}_j), i, j = 1, \dots, p, \quad (3)$$

где g – число классов; p – число дискриминантных переменных; n_k – число наблюдений в k -й группе; x_{ikm} , x_{jkm} – величина переменной i (j) для m -го наблюдения в k -й группе; \bar{x}_{ik} , \bar{x}_{jk} – средняя величина переменной i (j) в k -й группе; \bar{x}_i , \bar{x}_j – среднее значение переменной i (j) по всем группам.

Матрицы W и B содержат всю основную информацию о зависимости внутри групп и между группами. Нахождение коэффициентов дискриминантных функций сводится к решению задачи о собственных значениях и векторах:

$$\begin{aligned} \sum b_{1i} v_i &= \lambda \sum w_{1i} v_i, \\ \sum b_{2i} v_i &= \lambda \sum w_{2i} v_i, \\ \sum b_{pi} v_i &= \lambda \sum w_{pi} v_i, \end{aligned}$$

где λ – собственное число, v_i – собственный вектор.

Каждое решение, имеющее свое собственное значение λ и свою последовательность v_i , соответствует одной дискриминантной функции. Компоненты собственного вектора v_i можно использовать в качестве коэффициентов дискриминантной функции

$$\beta_i = v_i \sqrt{n-g}, \beta_0 = -\sum_{i=1}^p \beta_i x_i,$$

где n – общее число наблюдений по всем группам.

Общее число дискриминантных функций не превышает числа дискриминантных переменных и, по крайней мере, на единицу меньше числа групп. Существует несколько характеристик, позволяющих оценить полезность дискриминантной функции: коэффициент корреляции r_i

$$r_i = \sqrt{\frac{\lambda_i}{1 + \lambda_i}},$$

Λ – статистика Уилкса

$$\Lambda = \prod_{i=k+1}^g \left(\frac{1}{1 + \lambda_i} \right),$$

где k – число вычисленных функций.

Чем больше величина r_i , тем лучше раздельная способность дискриминантной функции. Чем меньше значение Λ , тем значимее соответствующая дискриминантная функция.

Новый объект относится к классу k , для которого значение дискриминантной функции d_{km} является максимальным.

При невыполнении требований, перечисленных в начале данного раздела, в использовании данного алгоритма могут возникать проблемы. Однако в работах по исследованию подлинности продуктов питания (классификация напитков и пищевого сырья по набору различных параметров и свойств) была показана достаточная устойчивость алгоритма дискриминантного анализа [126–128]. Так, LDA применяли для классификации 165 образцов французского и баскского сидра, каждый из которых характеризовали набором из 27 параметров, определенных высокоэффективной жидкостной хроматографией и указывающих на содержание различных полифенолов в образцах [12]. Случайным образом массив данных разбили на обучающую (75% образцов) и тестовую (25% образцов) выборки. Результаты использования метода линейного дискриминантного анализа

следующие: на этапе обучения доли правильно отнесенных образцов французского и баскского сидра составила 99 % и 100 %, а на этапе классификации тестируемых образцов – 96 % и 100%, соответственно.

Вместе с тем, известны и задачи, при решении которых линейный дискриминантный анализ привел к неудовлетворительным результатам (классификация образцов сыра с ошибкой 44% [129], образцов свинины с ошибкой 32% [106]).

Следует отметить, что LDA нецелесообразно применять, если границы кластеров перекрываются. Для решения таких задач перспективнее алгоритмы, основанные на теории нечетких множеств. Теория нечетких множеств предложена более 40 лет назад американским математиком Лотфи Заде. Методы на основе нечеткой логики позволяют работать с данными, которым свойственны неопределенность и неясность, и допускается возможность частичной принадлежности объектов к нескольким классам (один и тот же объект может одновременно относиться к различным классам, но с разной степенью принадлежности). Нечеткий линейный дискриминантный анализ (fuzzy linear discriminant analysis, FLDA) является усовершенствованным алгоритмом линейного дискриминантного анализа. Он способен предоставить больше информации о структуре изучаемых наборов данных [118, 130–132].

FLDA состоит из следующих шагов [133]:

1. Расчет нечетких принадлежностей каждого образца ко всем классам с помощью нечеткого алгоритма k-средних (fuzzy c-means, FCM).

FCM-алгоритм предполагает, что объекты принадлежат всем кластерам с определенной функцией принадлежности. Степень принадлежности определяется расстоянием от объекта до соответствующих кластерных центров (чаще всего используют евклидову метрику). Данный алгоритм итерационно вычисляет центры кластеров и новые степени принадлежности объектов. Алгоритм основан на минимизации целевой функции:

$$J = \sum_{k=1}^g \sum_{m=1}^{n_k} u_{mk}^l \sqrt{(x_{ikm} - c_k)^2},$$

где c_k – центр k -го класса, u_{mk} – степень принадлежности объекта m кластеру k , l – экспоненциальный вес – действительное число, большее 1. Задачей FCM-алгоритма является разбиение набора объектов на заданное число кластеров. Первоначальная матрица принадлежностей объектов к классам генерируется случайным образом. Затем выполняется расчет центров кластеров, евклидовых расстояний и новых значений принадлежностей объектов к классам:

$$c_k = \frac{\sum_{m=1}^{n_k} u_{mk}^l x_{ikm}}{\sum_{m=1}^{n_k} u_{mk}^l};$$

$$u_{mk} = \frac{1}{[(x_{ikm} - c_k)^2 \sum_{j=1}^c \frac{1}{(x_{imk} - c_k)^2}]^{\frac{2}{l-1}}}.$$

Работа алгоритма прекращается, когда квадрат разности между полученными значениями принадлежностей и значениями принадлежностей, полученных на предыдущей итерации, становится меньше заданного параметра точности $\|u - u^*\| < \varepsilon$ [130, 134, 135].

2. Расчет матриц W и B внутриклассовых и межклассовых вариаций. Показатели, входящие в состав формул расчета внутриклассовых и межклассовых вариаций (2, 3), вычисляются следующим образом:

$$\begin{aligned} n_k &= n_k u_{mk}, \\ x_{ikm} &= u_{km} x_{ikm}, \\ \bar{x}_{ik} &= \frac{\left(\sum_{m=1}^{n_k} u_{mk} x_{ikm} \right)}{n_k}, \\ \bar{x}_i &= \frac{\left(\sum_{k=1}^g \sum_{m=1}^{n_k} u_{mk} x_{ikm} \right)}{\sum_{k=1}^g n_k}. \end{aligned}$$

Аналогично вычисляются x_{jkm} , \bar{x}_{jk} , \bar{x}_j .

3. Отбор дискриминантных переменных и расчет дискриминантных функций.

На примере предсказания зрелости арахиса исследователи показали, что применение алгоритма FLDA несколько повышает эффективность классификации по сравнению с методом LDA [136]. При разбиении набора образцов арахиса по данным об их спектрах в ближней ИК-области на 3, 5 и 6 групп методом LDA точность классификации составила 42%, 56% и 70% соответственно, а при использовании алгоритма FLDA точность разделения повысилась до 45%, 63% и 73%.

Деревья классификации и регрессии (Classification and regression trees, CART)

Деревья регрессии и классификации, известные также под общим названием как деревья решений (Decision Tree, DT), являются популярным методом решения задач классификации и прогно-

зирования с обучением. После обучения классифицируемые наблюдения представляются в виде последовательностей предложений "If-Then" ("Если-То"), организованных в виде дерева.

Любое дерево решений, по существу, представляет собой древовидный граф. Эта структура данных состоит из узлов, соединенных друг с другом ребрами (ветвями). При этом не допускается, чтобы ребра образовывали цикл, так как в противном случае дерево превращается в граф, отличный от древовидного. Дерево имеет один особый узел – корневой (корень). Корневой узел является основой дерева, поскольку от него можно перейти по дереву к любому другому узлу. В конце любой цепочки подряд идущих ребер находятся листовые узлы (листья, терминальные узлы). Корень дерева содержит все классифицируемые данные, а листья – определенные классы после выполнения классификации. Промежуточные узлы дерева представляют пункты принятия решения о выборе или выполнения тестирующих процедур, а ветви, выходящие из узлов, соответствуют количеству возможных исходов тестирующей процедуры. Узел, который можно разбить на два новых, называется родительским (узел-родитель), а новые узлы – узлами-потомками (подчиненными).

Для построения дерева решений алгоритм сначала создает корневой узел. В алгоритме CART каждый узел дерева решений имеет двух потомков. На каждом шаге построения дерева решений правило, формируемое в узле, делит обучающую выборку на две части – часть, в которой выполняется правило (потомок-right), и часть, в которой правило не выполняется (потомок-left). В каждом узле разбиение идет только по одной переменной. Общее правило для выбора переменной можно сформулировать следующим образом: выбранная переменная должна разбить выборку данных так, чтобы получаемые в итоге подмножества состояли из объектов, принадлежащих к одному классу (или были максимально приближены к этому, т.е. количество объектов из других классов в каждом из этих подмножеств должно быть как можно меньшим). На каждом шаге построения дерева алгоритм последовательно сравнивает все возможные разбиения для всех переменных и выбирает наилучшее разбиение.

Построение дерева решений относится к классу обучения с учителем. Качество разбиения оценивается по следующей формуле:

$$G = \frac{1}{N} \left[L \left(1 - \frac{1}{L^2} \sum_{i=1}^n l_i^2 \right) + R \left(1 - \frac{1}{R^2} \sum_{i=1}^n r_i^2 \right) \right],$$

где N – число образцов в узле-родителе; L и R – число образцов в левом и правом узле-потомке

соответственно; l_i и r_i – число объектов i -го класса в левом/правом узле-потомке.

Лучшим будет то разбиение, для которого величина G максимальна.

Недостатком метода является то, что дерево может оказаться переученным. Поэтому следующий этап алгоритма CART – отсечение (сокращение, уменьшение) дерева (minimal cost-complexity tree pruning). Основная проблема отсечения – большое число возможных отсеченных поддеревьев для одного дерева. Обозначим через T число терминальных узлов дерева, а через $R(T)$ – ошибку классификации дерева, равную отношению числа неправильно классифицированных объектов к числу объектов в обучающей выборке. Определим $C_\alpha(T)$ – полную стоимость (показатель затраты-сложность) дерева T как

$$C_\alpha(T) = R(T) + \alpha T,$$

где α – параметр, изменяющийся от 0 до $+\infty$.

Полная стоимость дерева включает две компоненты – ошибку классификации и штраф за его сложность. Для сокращения необходимо проверить пары узлов, имеющих общего родителя, и выяснить, насколько увеличится $C_\alpha(T)$ при их объединении. Если увеличение окажется меньше заданного порога, то оба листовых узла сливаются в один, для которого множество результатов получается объединением результатов в исходных листьях. Идея заключается в том, чтобы вычислить значения α для каждого узла t в дереве T и выбрать "слабые связи", т.е. узлы t , для которых величина

$$\alpha(t) = \frac{R(t) - R(T)}{T - 1}$$

является наименьшей.

Так избегают феномена переучивания.

Построение дерева регрессии во многом схоже с построением дерева классификации. Сначала строится дерево максимального размера, которое затем обрезается до оптимального размера. Процесс построения дерева происходит последовательно. На первом шаге получаем регрессионную оценку как константу по всему пространству объектов. Константу вычисляем как среднее арифметическое выходной переменной в обучающей выборке. Если обозначить все значения выходной переменной как Y_1, Y_2, \dots, Y_n , то регрессионную оценку находят как

$$\hat{f}(x) = \left(\frac{1}{n} \sum_{i=1}^n Y_i \right) I_R(x),$$

где R – пространство обучающих объектов, n – число объектов, $I_R(x)$ – индикаторная функция пространства (набор правил, описывающих попадание переменной x в пространство).

На втором шаге пространство делят на две части. Регрессионная оценка принимает вид

$$\hat{f}(x) = \left(\frac{1}{I_1} \sum_{I_1} Y_i \right) I_{R1}(x) + \left(\frac{1}{I_2} \sum_{I_2} Y_i \right) I_{R2}(x)$$

Оценкой лучшего разбиения служит сумма квадратов разностей

$$E = \sum_{i=1}^n (Y_i - \hat{f}(x_i))^2,$$

которую нужно минимизировать. Процесс разбиения продолжается до тех пор, пока сумма квадратов разностей не станет меньше некоторого заданного порога.

Отсечение и выбор финального дерева производится по процедурам, аналогичным используемым для дерева классификации. Единственное отличие состоит в оценивании ошибки ответа дерева, которая рассчитывается по формуле:

$$R(\hat{f}) = \frac{1}{n} \sum_{i=1}^n (Y_i - \hat{f}(x_i))^2.$$

Таблица 6. Результаты применения алгоритма CART для классификации образцов пшеницы, сливок, зеленого чая по данным ИК-спектроскопии

Объект	Обучающая выборка	Тестовая выборка	Ненадежность классификации, %
	(число образцов)		
Пшеница	70	30	13
Сливки	450	537	5
Зеленый чай	70	50	18

Здесь, а также в табл. 7 и в табл. 11, оценкой ненадежности служит ошибка при классификации образцов тестовой выборки.

Метод опорных векторов (Support vector machines, SVM)

Метод опорных векторов – группа алгоритмов, основанных на обучении с учителем, – используется для решения задач бинарной классификации и регрессионного анализа. В основе метода лежит понятие плоскостей решений, определяющих границы принятия решений, т.е. разделение объектов в пространстве переменных осуществляется с помощью построения гиперплоскости [98, 146–148].

Гиперплоскость строится на основании обучающей выборки – множества объектов x_i , заданных в виде векторов переменных с явным ука-

Стоимость дерева равна:

$$C_\alpha(\hat{f}) = R(\hat{f}) + \alpha \hat{f}.$$

После создания дерева решений появляется возможность классифицировать неизвестные выборки данных. Определение класса осуществляется путем перехода по дереву, начиная от корневого узла, в направлении к листовым узлам [137–142].

К достоинствам метода построений деревьев регрессии и классификации относятся их иерархическое строение и возможность последовательно изучать эффект влияния отдельных переменных.

Алгоритм CART применяли для исследования образцов воды [143], подлинности продуктов вина [144], спирта [145], пшеницы, сливок, зеленого чая [139].

Рассмотрим пример использования алгоритма. Для 100 образцов пшеницы, 987 образцов сливок и 120 образцов зеленого чая получили ИК-спектры и разделили, применяя алгоритм CART, каждый из полученных массивов данных на два класса [139].

В табл. 6 представлены оценки ненадежности работы алгоритма как доли неправильно классифицированных образцов тестовой выборки к общему числу образцов в данной выборке.

занием принадлежности к одному из классов c_i (+1, –1). Гиперплоскость строится таким образом, чтобы максимизировать ширину границы (ширину полосы) между положительной и отрицательной частями обучающей выборки.

Таким образом, в построении гиперплоскости участвуют только опорные вектора (support vectors), т.е. объекты на границе между частями обучающей выборки [149, 150].

Гиперплоскость можно представить в виде

$$w \cdot x = b.$$

Необходимо найти вектор w такой, что для некоторого граничного значения b и новой точки x_i выполняется условие

$$w \cdot x_i > b \Rightarrow c_i = 1;$$

$$w \cdot x_i < b \Rightarrow c_i = -1.$$

Если скалярное произведение вектора w на x_i больше некоторого порогового значения b , то новую точку относят к первой категории, если меньше – ко второй. Вектор w перпендикулярен искомой разделяющей прямой, а значение b зависит от кратчайшего расстояния между разделяющей прямой и началом координат [151, 152].

Если обучающая выборка линейно разделима, то выбрать гиперплоскости можно таким образом, чтобы между ними не лежала ни одна точка обучающей выборки, а затем максимизировать расстояние между гиперплоскостями. Ширина полосы между ними равна

$$\left\langle \frac{x_+ - x_-}{\|w\|}, \frac{w}{\|w\|} \right\rangle = \frac{\langle w, x_+ \rangle - \langle w, x_- \rangle}{\|w\|^2} = \frac{(w_0 + 1) - (w_0 - 1)}{\|w\|^2} = \frac{2}{\|w\|^2}.$$

Ширина полосы максимальна, когда норма вектора w минимальна.

Для исключения всех точек из полосы необходимо убедиться, что для всех i ($1 \leq i \leq n$) выполняется условие

$$c_i(w \cdot x_i - b) \geq 1 \quad (4)$$

Оптимальную разделяющую гиперплоскость строят, минимизируя $\|w\|$ с учетом условия (4) [152, 153].

При работе с многопараметрическими данными рекомендуют определить наиболее информативные переменные. Их выбор основывается на подсчете значения F для каждой переменной:

$$F(i) = \frac{(x_i^{(+)} - \bar{x}_i)^2 + (x_i^{(-)} - \bar{x}_i)^2}{\frac{1}{n_+ - 1} \sum_{k=1}^{n_+} (x_{k,i}^{(+)} - \bar{x}_i^{(+)})^2 + \frac{1}{n_- - 1} \sum_{k=1}^{n_-} (x_{k,i}^{(-)} - \bar{x}_i^{(-)})^2},$$

где n_+ – количество положительных образцов, n_- – количество отрицательных образцов, $x_i^{(+)}, x_i^{(-)}, \bar{x}_i$ – средние значения i -го признака в положительной, отрицательной частях и всей выборке, $x_{k,i}^{(+)}, x_{k,i}^{(-)}$ – значения i -го признака k -го положительного и отрицательного объектов, соответственно.

Переменные упорядочивают по значению F , причем переменные с наименьшими значениями F отбрасывают [154].

В случае линейно разделимой выборки решение задачи квадратичной оптимизации имеет вид

$$w = \sum_{i=1}^n \lambda_i c_i x_i,$$

где λ_i – вектор двойственных переменных. Если $\lambda_i > 0$, то объект обучающей выборки x_i называется опорным вектором [150, 155].

Однако большинство задач классификации не так просты (например, в случае перекрывания классов), и часто необходимо строить оптимальные разбиения множеств объектов с помощью разделителей намного более сложной структуры, чем линейный разделитель. В этом случае исходные объекты переупорядочивают при помощи специального класса математических функций, называемых ядрами. Этот процесс переупорядочивания называют еще преобразованием объектов (перегруппировкой). В этом случае классифицирующая функция имеет вид [154, 156]

$$g(x) = \sum_{i=1}^n \lambda_i c_i K(x_i, x) + b,$$

где λ_i – коэффициенты, определяемые в ходе оптимизации, x_i – опорные векторы, K – ядерная функция (kernel function), которая используется для перехода к нелинейной разделяющей поверхности.

Наибольшее распространение получили следующие ядра [153, 157]:

– полиномиальное $K(x_i, x) = (\langle x_i, x \rangle + 1)^d$;

– радиальная базисная функция

$K(x_i, x) = \exp(-\beta \|x_i - x\|^2)$, где β – параметр;

– радиальная базисная функция Гаусса

$K(x_i, x) = \exp\left(-\frac{\|x_i - x\|^2}{2\sigma^2}\right)$, где σ^2 – дисперсия.

Достоинства метода SVM следующие: задача квадратичной оптимизации имеет единственное решение; правила классификации выбираются не с помощью некоторых эвристических предположений, при оптимизации некоторой функции; линейный классификатор используется для работы с нелинейно разделяемыми данными; алгоритм сочетает простоту использования с эффективностью; есть возможность при использовании различных ядер реализовать другие методы классификации (нейронные сети с сигмоидными функциями активации, нейронные сети с радиальными функциями активации) [151, 153].

Известны успешные примеры использования метода опорных векторов для исследования каче-

ства образцов мяса [158], яблок [159], пшеницы [160], сыра [161].

Рассмотрим применение метода опорных векторов для классификации образцов сыра.

Набор данных состоит из 60 образцов сыра 6 различных видов, по 10 образцов каждого вида [161]. Каждый из образцов сыра описывали 35 физическими, химическими характеристиками и характеристиками внешнего вида. Обучающая выборка включала 48 образцов, тестовая – 12.

Классификацию проводили с использованием кросс-валидации (перекрестной проверки достоверности), т.е. на каждом шаге кросс-валидации формировали обучающую и тестовую выборки, а ненадежность метода оценивали как среднее всех ошибок классификации. В качестве ядра использовали радиальную базисную функцию.

В табл. 7 приведены оценки ненадежности метода опорных векторов для классификации образцов 6 различных видов сыра.

Таблица 7. Результаты классификации образцов сыра методом опорных векторов (%)

Вид 1	Вид 2	Вид 3	Вид 4	Вид 5	Вид 6	Ненадежность классификации, %
96	76	66	74	100	86	17

Проекция на латентные (скрытые) структуры (Projection on latent structures, PLS) и PLS-регрессия

Существуют две трактовки фундаментального метода PLS: Partial Least Squares – "частичные" или "частные наименьшие квадраты" и Projection on Latent Structures, что дословно переводится как "проекция на латентные структуры". Рекомендуется использовать вторую трактовку, т.к. она отражает суть метода [94, 109].

PLS – популярный метод многомерной градуировки, мощная альтернатива методу главных компонент; PLS-анализ предоставляет хорошую возможность интерпретации результатов, используя при этом меньшее число главных компонент (в данном случае PLS-компонент) [104, 109, 162].

В отличие от метода главных компонент, в методе PLS проводится одновременная декомпозиция (разложение, разбиение) двух матриц: X (предикторы, независимые переменные, например, спектральные полосы) и Y (отклики, наблюдения, зависимые переменные, например, концентрации компонента):

$$X = TP' + E = t_1 p_1' + t_2 p_2' + \dots + t_n p_n' + E,$$

$$Y = UQ' + F = u_1 q_1' + u_2 q_2' + \dots + u_n q_n' + F,$$

где T и U представляют матрицы счетов, P и Q – матрицы нагрузок, E и F – матрицы остатков. Существует две популярные разновидности метода проекций на латентные структуры: PLS1 и PLS2. Модель PLS1 строится для единственной переменной y , например, для концентрации одного компонента смеси, PLS2 рассчитывается для нескольких переменных Y одновременно, т.е. позволяет моделировать любую комбинацию переменных совместно. Соответственно отличаются и расчетные алгоритмы этих методов (табл. 8, 9), на каждой итерации вы-

числяется одна PLS-компонента. Как правило, методу PLS предшествует центрирование или нормирование исходных матриц X и Y (см. раздел о методе главных компонент) [163].

Разложения матриц X и Y методом PLS2 тесно связаны друг с другом. Первый набор значений векторов-нагрузок p_1 и q_1 получают из условия максимизации ковариации (меры линейной зависимости) между соответствующими векторами X и Y . Проецированием данных Y на q_1 получают первый набор значений вектора счетов u_1 . Матрицы X и Y косвенно связаны через векторы-счета: вектор счетов u_1 является отправным пунктом для расчета вектора t_1 -счетов. При этом в методе PLS2 начальный вектор t_1 заменяется на u_1 , и, таким образом, структура данных Y напрямую влияет на декомпозицию матрицы X . Вектор u_1 (отражающий структуру Y), при помощи которого выбирается первое направление в декомпозиции матрицы X , дает новые нагрузки для матрицы X , обозначаемые символом w (loadings-weights – взвешенные нагрузки).

После этого рассчитываются новые t -векторы пространства X с использованием векторов нагрузок w . Затем эти t -векторы применяются как стартовые, заменяющие u_1 . Таким образом, структура данных X также влияет на декомпозицию Y . Итерационная замена независимых векторов $u_i \rightarrow t_i$ и $t_i \rightarrow u_i$ (обмен векторами счетов) осуществляется до достижения сходимости. Вектор w_1 определяет направление первой PLS-компоненты в пространстве X , на которое проецируются все образцы. Обычно это направление

Таблица 8. Алгоритм PLS1

№ шага	Процедура	Пояснение
1	$w_i = \frac{X_i^t y_i}{ X_i^t y_i }$	Расчет нормализованного вектора взвешенных нагрузок w
2	$t_i = X_i^t w_i$	Расчет вектора весов t
3	$q_i = \frac{y_i^t t_i}{ t_i^t t_i }$	Расчет нагрузки q переменной i
4	$p_i = \frac{X_i^t t_i}{t_i^t t_i}$	Расчет вектора нагрузок p
5	$E_i = X_i - T_i^t P_i$ $F_i = y_i - b_i T_i Q_i^t$	Расчет остатков
6	$i = i + 1$	Переход к следующей PLS-компоненте

Таблица 9. Алгоритм PLS2

№ шага	Процедура	Пояснение
1	u_i	Выбор начального приближения u
2	$w_i = \frac{X_i^t u_i}{ X_i^t u_i }$	Расчет нормализованного вектора взвешенных нагрузок w
3	$t_i = X_i^t w_i$	Расчет вектора счетов t
4	$q_i = \frac{Y_i^t t_i}{ Y_i^t t_i }$	Расчет нормализованного вектора нагрузок q
5	$u_i = Y_i^t q_i$	Расчет вектора счетов u
6	$ t_{i,new} - t_{i,old} < \lim$	Проверка сходимости алгоритма: если условие не выполняется – переход к шагу 2.
7	$p_i = \frac{X_i^t t_i}{t_i^t t_i}$	Расчет вектора нагрузок p
8	$b_i = \frac{u_i^t t_i}{t_i^t t_i}$	Расчет внутреннего коэффициента регрессии
9	$E_i = X_i - T_i^t P_i$ $F_i = Y_i - b_i T_i Q_i^t$	Расчет остатков
10	$i = i + 1$	Переход к следующей PLS-компоненте

не совпадает с направлением p_1 , так как в PLS-алгоритме одновременно проводится и максимизация ковариации (t, u) . Различие между направлениями этих альтернативных компонент показывает, насколько сильно матрица Y повлияла на разбиение матрицы X [162, 164, 165]. Важность

обеих матриц P и W становится очевидной, если представить формулу расчета коэффициента B формального регрессионного уравнения $Y = XB$ [162]:

$$B = W(P^t W)^{-1} Q^t .$$

Значения B часто используются в практических целях при классификации новых объектов:

$$\hat{Y} = X_{new}B.$$

Алгоритм PLS1 проще, чем PLS2: пропадает итерационная подмена $u_i \rightarrow t_i$ и $t_i \rightarrow u_i$.

Интерпретация PLS-моделей служит для изучения внутренней структуры данных: определения групп объектов, выбросов (график $t-u$), взаимосвязей между переменными (графики $w-q, p-w$). Методы многомерной калибровки также позволяют заменить прямое измерение интересующего нас свойства измерением другого свойства, коррелирующего с первым [159].

Алгоритм PLS часто используется для решения задач, связанных с изучением качества и происхождения продуктов питания и пищевого сырья. С его помощью (как и с помощью метода главных компонент) выявляют взаимосвязи данных в больших массивах результатов измерений, полученных с помощью ИК-спектроскопии с Фурье-преобразованием [68, 166, 167], спектроскопии в ближнем ИК-диапазоне [168, 169], Рамановской спектроскопии [170], спектроскопии в среднем ИК-диапазоне [171, 172].

В качестве примера рассмотрим применение алгоритма проекций на латентные структуры при исследовании возможности идентифицировать яблочные соки [172]. Для этого были созданы лабораторные стандарты с различным содержанием яблочного сока: 2%, 4%, 6%, 8%, 10%, 16%, 20%, 25%, 50%, 70% и 100%. Набор стандартных образцов разделили на две группы, для каждой группы определили обучающую и тестовую вы-

борки. Первая группа включала образцы с содержанием яблочного сока от 2 до 20%; общее число образцов 173, 134 из которых составили обучающую выборку, а 39 образцов использовали для проверки алгоритма. Вторая группа включала образцы с содержанием яблочного сока от 25 до 100%; общее число образцов 130, 86 из которых составили обучающую выборку и 44 – тестовую. Для каждого стандартного образца получили инфракрасные спектры в среднем ИК-диапазоне.

К полученному массиву результатов измерений применили метод PLS. При этом матрица X – это набор полученных спектральных полос, матрица Y – стандартные образцы с различным содержанием сока. Матрица Y представляет собой двоичную кодировку этих образцов. В результате работы выделили две латентные компоненты. Полученную модель применили для классификации 23 коммерческих образцов сока: 2 образца принадлежали к первой группе объектов, 21 образец – ко второй группе. В табл. 10 приведены два параметра, характеризующих работу алгоритма PLS: эффективность (отношение правильно отнесенных образцов в соответствующий класс к общему числу образцов, принадлежащих этому классу, %) и чувствительность (отношение числа образцов, правильно отнесенных в соответствующий класс, к общему числу образцов, классифицированных в этот класс в результате работы алгоритма, %). В результате применения алгоритма PLS к промышленным образцам сока только один образец второй группы был идентифицирован неправильно.

Таблица 10. Значения эффективности и чувствительности (%) метода PLS при его применении для идентификации стандартных образцов с различным содержанием яблочного сока

Содержание яблочного сока в образцах	2%	4%	6%	8%	10%	16%	20%	25%	50%	70%	100%
Эффективность	100	100	100	50	56	60	63	81	92	100	100
Чувствительность	100	100	85	33	83	50	83	94	100	100	100

Очень мощным инструментом выявления фальсифицированных продуктов питания является метод дискриминантного анализа с помощью регрессии на латентные структуры (PLS discriminant analysis, PLS-DA). Его используют для работы с большими массивами данных, полученных в результате применения, например, масс-спектрометрии [173, 174], спектроскопии в ближнем ИК-диапазоне и видимой спектроскопии [131, 175], Рамановской спектроскопии [70]. Идея этого подхода состоит в том, что правила дискримина-

ции для K классов задаются линейными регрессионными уравнениями вида

$$XB = D,$$

где X – полная матрица всех исходных данных ($I \times J$), B – матрица неизвестных коэффициентов ($J \times K$), а D – это специальная матрица ($I \times K$), которая состоит из нулей и единиц. При построении матрицы D единицы ставят только в те строки (строки соответствуют исследуемым объектам), которые принадлежат классу, соответ-

ствующему номеру столбца. Регрессионная задача решается методом PLS, что позволяет в дальнейшем применять построенную регрессию для предсказания принадлежности новых объектов. Для этого строят прогноз отклика нового объекта и сравнивают результат с нулем или единицей [176]. На примере классификации образцов женьшеня [70] покажем эффективность использования алгоритма PLS-DA. Были отобраны 50 китайских и корейских образцов женьшеня из различных регионов этих стран. Все образцы высушили и измельчили до порошкообразного состояния и для каждого образца получили ИК-спектры в ближней

ИК-области и Рамановские спектры. Весь полученный массив данных случайным образом разделили на обучающую выборку для нахождения оптимальных параметров классификации и тестовую выборку для проверки созданной модели.

Перед применением метода PLS-DA для ИК-спектров вычислили вторые производные для улучшения разрешения (для точного определения положения перекрывающихся спектральных полос). В табл. 11 представлена информация об использованном массиве данных и результатах работы метода PLS-DA.

Таблица 11. Сведения о спектральных характеристиках образцов женьшеня и результаты работы PLS-DA-анализа

Параметр	ИК-спектроскопия	Рамановская спектроскопия
Число образцов в обучающей выборке (корейские/китайские)	70 (35/35)	70 (35/35)
Число образцов в тестовой выборке (корейские/китайские)	30 (15/15)	30 (15/15)
Диапазон измерений	400–2500 нм	250–1700 см ⁻¹
Число PLS-компонент	3	6
Число неправильно классифицируемых образцов в обучающей/тестовой выборке	0/0	5/3
Ненадежность классификации, %	0	9

Искусственные нейронные сети (Artificial neural networks, ANN)

Нейронные сети – это математические модели, современные вычислительные системы, состоящие из элементарных единиц обработки информации (нейронов), накапливающих экспериментальные знания и предоставляющих их для последующей обработки. Нейронная сеть сходна с мозгом с двух точек зрения:

– знания поступают в нейронную сеть из окружающей среды и используются в процессе обучения;

– для накопления знаний применяются связи между нейронами, называемые синаптическими весами.

В литературе нейронные сети часто называют нейрокомпьютерами, сетями связей (connectionist network), интеллектуальными системами и т.д. [177–181]. Нейрон является составной частью нейронной сети. Функционирование нейрона определяется формулами:

$$NET = \sum_i w_i x_i,$$

$$OUT = F(NET - \theta),$$

где x_i – входные сигналы (исходные данные),

w_i – весовые коэффициенты нейронов, NET –

взвешенная сумма входных сигналов, θ – пороговый уровень данного нейрона, F – функция активации (линейная, логистическая, экспоненциальная и т.д.), OUT – выходной сигнал нейрона.

Выполняемые сетью функции можно разделить на несколько основных групп: аппроксимации и интерполяции; распознавания и классификации образов; сжатия данных; прогнозирования; идентификации; управления; ассоциации. Важнейшее свойство нейронных сетей, свидетельствующее об их огромном потенциале и широких практических возможностях, состоит в параллельной обработке информации одновременно всеми нейронами. Благодаря этой способности при большом количестве межнейронных связей достигается значительное ускорение процесса обработки информации. Другое не менее важное свойство нейронной сети – способность к обучению и обобщению полученных знаний. Сеть обладает чертами искусственного интеллекта. Натренированная на ограниченном множестве обучающих выборок, она обобщает накопленную информацию и вырабатывает ожидаемую реакцию применительно к данным, не обработанным в процессе обучения. Процедура, используемая для процесса обучения, называется алгоритмом обучения (learning algorithm). Обучение осуществляется путем последовательного предъявления входных векторов

с одновременной подстройкой весов в соответствии с определенной структурой. В процессе обучения веса сети постепенно становятся такими, чтобы каждый входной вектор вырабатывал выходной вектор. Различают алгоритмы обучения с учителем и без учителя.

Обучение с учителем предполагает, что для каждого входного вектора существует целевой вектор, представляющий собой требуемый выход. Обучение без учителя не нуждается в целевом векторе для выходов.

Обучающий алгоритм подстраивает веса сети так, чтобы получались согласованные выходные векторы, т.е. чтобы предъявление достаточно близких входных векторов давало одинаковые выходы. Существуют и специфические алгоритмы обучения, например, Хебба, Кохонена, Розенблатта, Коши, процедура обратного распространения ошибки [181–186].

Обучаемость, способность к обобщению, к абстрагированию, параллельность обработки информации способствуют применению искусственных нейронных сетей в исследовании подлинности и фальсификации продуктов питания. В табл. 12 приведены основные парадигмы искусственных нейронных сетей [182, 187, 188],

используемые для идентификации продуктов питания и напитков.

Искусственные нейронные сети находят свое применение во многих областях аналитической химии, в особенности при контроле качества продуктов питания [189]. В качестве примера использования алгоритмов искусственных нейронных сетей рассмотрим применение сети радиального основания (Radial Basis Function Network) для классификации образцов пищевого масла (оливкового масла первого прессования, оливкового масла не первого прессования, подсолнечного масла, рапсового масла) [81]. Общее число образцов масла 480, из которых 320 составили обучающую выборку, 160 образцов – тестовую. Для каждого образца для получения экспериментального массива данных применили люминесцентную спектроскопию.

Структура RBF-сети следующая: 3 входных нейрона, 48 скрытых (промежуточных), 16 выходных нейронов. Число правильно классифицируемых образцов пищевого масла в результате применения нейронной сети радиального основания равнялось 476, т.е. лишь 4 образца были идентифицированы неправильно, ненадежность работы алгоритма составила 1%.

Таблица 12. Основные парадигмы нейронных сетей

Название	Область применения	Недостатки	Преимущества	Примеры использования
Нечеткая сеть теории адаптивного резонанса (Fuzzy Adaptive Resonance Theory Network)	Распознавание образов, кластерный анализ	Неограниченное увеличение числа нейронов, проблемы при наличии в данных шумов	Обучение без учителя	Классификация образцов кофе [190]
Сеть встречного распространения (Counter Propagation Network)	Распознавание и восстановление образов (ассоциативная память), сжатие данных, статистический анализ	Сеть не дает возможности построить точные аппроксимации	Простота алгоритма. Сеть дает возможность извлекать статистические характеристики	Идентификация геологического происхождения образцов минеральных вод [191]
Сеть радиального основания (Radial Basis Function Network)	Распознавание образов, классификация	Заранее должно быть известно число эталонов	Отсутствие этапа обучения	Определение содержания меланина в образцах молока [192]
Генетический алгоритм (Genetic Algorithm)	Распознавание образов, классификация, прогнозирование	Сложны для понимания и программной реализации	Высокая скорость обучения, сокращение общего времени обучения	Определение содержания сахара в шоколаде [193]
Сеть Хемминга (Hamming Network)	Распознавание образов, классификация, ассоциативная память	Применение только бинарных входных сигналов, работа с слабо зашумленными сигналами	Сеть работает предельно быстро и просто, используется легкий алгоритм формирования весов и смещений сети	Классификация растительных масел [194]
Сеть Хопфилда (Hopfield Network)	Ассоциативная память, задачи оптимизации, распознавание образов	Сеть обладает небольшой емкостью	Позволяет восстановить искаженные сигналы	Определение качества воды [195]
Сеть Кохонена (Kohonen's Neural Network)	Кластерный анализ, классификация, распознавание образов	Сеть может быть использована для кластерного анализа только в случае, если заранее известно число кластеров	Сеть способна функционировать в условиях помех	Исследование образцов голубого сыра [196]

ЗАКЛЮЧЕНИЕ

Расширение круга задач и развитие методологии качественного химического анализа приводит к его пониманию как совокупности экспериментальных и расчетных процедур, обеспечивающих классификацию (в частности, идентификацию) объектов по их химическим, физико-химическим и другим характеристикам. Возрастает роль хемометрических методов, без которых невозможна обработка больших массивов многомерных данных. В связи с этим, добиваясь высокой надежности выводов качественного анализа, химик-аналитик должен внимательно выбирать не только экспериментальные методы, но и хемометрические средства обработки многомерных данных.

При решении задач, связанных с контролем качества и подлинности продуктов питания и пищевого сырья, получили широкое распространение как традиционные, так и достаточно новые хемометрические методы. Метод главных компонент служит для предварительной обработки и извлечения информации из многомерных массивов данных, позволяя существенно снизить размерность массива данных при минимальной потере информации. Метод формального независимого моделирования аналогий классов, позволяющий относить образцы к одному или нескольким классам или к ни одному из смоделированных классов, особенно удачно использовался для обработки спектральных данных. Дискриминантный анализ, несмотря на широкое распространение, далеко не всегда обеспечивает высокую надежность аналитических выводов. Деревья классификации и регрессии – популярный метод решения задач классификации – привлекателен тем, что позволяет выявлять влияние тех или иных переменных на результат работы алгоритма. Метод опорных векторов, популярность которого растет особенно быстро, эффективно решает задачи бинарной классификации. Проекция на латентные структуры – полезный метод многомерной градуировки, обеспечивающий выявление внутренней структуры данных и успешно применявшийся для обработки массивов спектральных данных. Искусственные нейронные сети – перспективная технология решения задач классификации, потенциал которой в исследовании качества и подлинности продуктов питания и сырья используется пока еще весьма ограниченно.

Представленные в статье примеры показывают, что различные хемометрические методы позволяют достаточно надежно устанавливать такие характеристики пищевых продуктов и сырья, как географическое происхождение, сорт, содержание различных компонентов и примесей. Можно ожидать, что алгоритмы классификации, распознава-

ния образов, искусственных нейронных сетей, нечеткой логики, эффективность которых уже установлена, будут все более широко использоваться в аналитической практике для проверки подлинности продуктов питания и пищевого сырья.

Ключевой проблемой является оценка надежности/ненадежности логических заключений, к которым приходит аналитик по результатам качественного анализа.

До настоящего времени господствует статистический (эвристический) подход, сводящийся к оценке ненадежности классификации по доле ошибочно классифицированных образцов в тестовой выборке. Ненадежность оценивают с помощью или стандартных алгоритмов, или алгоритмов, параметры которых были отрегулированы при обработке обучающей выборки. Поэтому, выбирая метод обработки массивов экспериментальных данных, предпочтение, по всей вероятности, следует отдавать классификации с обучением. Повысить уверенность в надежности классификации позволит и более широкое использование перекрестной оценки достоверности (кросс-валидации).

Вместе с тем, статистическая ("ad hoc") процедура, привлекая простотой и давая разумную оценку ненадежности решения данной классификационной задачи, имеет и существенный устранимый недостаток: трудно предвидеть, как поведет себя хемометрический алгоритм при переходе к обработке нового массива экспериментальных данных с, возможно, другой структурой и статистическими характеристиками.

В связи с постоянным расширением арсенала хемометрических методов³ актуальным становится вопрос: возможна ли, а если да, то какой должна быть неэвристическая (априорная, на основе вероятностных моделей экспериментальных данных) стратегия оценки ненадежности выводов, сделанных по результатам обработки многомерных массивов экспериментальных данных?

Как следует менять параметры алгоритмов, чтобы обеспечить надежность выводов при обработке с их помощью данных, обладающих разной структурой и статистическими свойствами?

То, что априорный подход определения ненадежности идентификации по хроматографическим и спектральным данным может быть весьма про-

³ О.Е. Родионова и А.Л. Померанцев указывают [94, С. 5]: "...Хемометрики изобретают все новые и новые методы... Делают они это так быстро, что математики ... не успевают не только раскритиковать их за это, но и просто понять, что же происходит в этой хемометрике".

дуктивен, убедительно показано в известной монографии [5].

Можно надеяться, что распространение методологии априорного (вероятностного) анализа надежности аналитических выводов на более широкую область окажется и возможным, и полезным.

Перспективными задачами исследований также можно считать выявление среди уже разработанных хемотрических алгоритмов наиболее эффективных для решения конкретных аналитических задач, а также разработку новых и модерни-

зацию существующих алгоритмов, позволяющих снизить требования к объему и точности исходных экспериментальных данных (устойчивых к наличию шума, пропусков, промахов и т.д.).

Авторы глубоко благодарны доктору химических наук, профессору, заведующему кафедрой аналитической химии Омского государственного университета им. Ф.М. Достоевского В.И. Вершинину за интерес к работе и ценные советы.

ЛИТЕРАТУРА

1. Trends Anal. Chem. – 2005. – 24, №6.
2. Milman B.L. Identification of chemical compounds // Trends Anal. Chem. – 2005. – 24, № 6. – P. 493–508.
3. Cárdenas S., Valcárcel M. Analytical features in qualitative analysis // Trends Anal. Chem. – 2005. – 24, № 6. – P. 477–487.
4. Попов А.А. Идентификация // Химическая энциклопедия. Т. 2. – М.: Сов. энциклопедия. – 1990. – С. 346–347.
5. Вершинин В.И., Дерендяев Б.Г., Лебедев К.С. Компьютерная идентификация органических соединений. – М.: Академкнига, 2002. – 197 с.
6. Hardcastle W.A. (Editor). Qualitative analysis: a guide to best practice. – Cambridge: Royal Society of Chemistry. 1998. – 24 p. <http://www.rsc.org/publishing/ebooks/1998/9780854044627.asp>
7. EUROCHEM/CITAC Guide "Quantifying uncertainty in analytical measurements", 2nd ed. – 2000. – 120 p.; Руководство ЕВРАХИМ/СИТАК "Количественное описание неопределенности в аналитических измерениях", 2-е изд. / Пер. с англ. Р.Л. Кадиса, Г.Р. Нежиховского, В.Б. Симины под ред. Л.А. Конопелько. – СПб, 2002. – 141 с.
8. Методические указания по оценке подлинности и выявлению фальсификации молочной продукции. Методические указания МУ 4.1./4.2.2484-09 – М.: Федеральный Центр гигиены и эпидемиологии Роспотребнадзора, 2009. – 26 с.
9. Kuchmenko T.A., Lisitskaya R.P., Golovanova V.A., Arsenova M.S. Detection of synthetic components in food matrices using piezoelectric resonators // Anal. Chem. – 2009. – 64, № 4. – P. 338–345.
10. Ivleva N.P., Niessner R., Panne U. Characterisation and discriminant of pollen by Raman microscopy // Anal. Bioanal. Chem. – 2005. – 381. – P. 261–267.
11. Coetzee P.P., Vanhaecke F. Classifying wine according to geographical origin via quadrupole-based ICP-mass spectrometry measurements of boron isotope ratios // Anal. Bioanal. Chem. – 2005. – 383. – P. 977–984.
12. Alonso-Salces R.M., Guyot S., Herrero C., Berrueta L.A., Drilleau J.F., Gallo B., Vicente F. Chemometrics characterisation of Basque and French ciders according to their polyphenolic profiles // Anal. Bioanal. Chem. – 2004. – 379. – P. 464–475.
13. Bellorini S., Strathmann S., Baeten V., Fumiere O., Berben G., Tirendi S., von Holst C. Discriminating animal fats and their origins: assessing the potentials of Fourier transform infrared spectroscopy, gas chromatography, immunoassay and polymerase chain reaction techniques // Anal. Bioanal. Chem. – 2005. – 382. – P. 1073–1083.
14. Poulli K.I., Mousdis G.A., Georgiou C.A. Synchronous fluorescence spectroscopy for quantitative determination of virgin olive oil adulteration with sunflower oil // Anal. Bioanal. Chem. – 2006. – 386. – P. 1571–1575.
15. Baeten V., von Holst C., Garrido A., Vancutsem J., Renier A.M., Dardenne P. Detection of banned meat and bone meal in feedstuffs by near-infrared microscopic analysis of the dense sediment fraction // Anal. Bioanal. Chem. – 2005. – 382. – P. 149–157.
16. Husted S., Mikkelsen B.F., Jensen J., Nielsen N.E. Elemental fingerprints analysis of barley (*Hordeum vulgare*) using inductively coupled plasma mass spectrometry, isotope-ratio mass spectrometry,

- and multivariate statistics // *Anal. Bioanal. Chem.* – 2004. – 378. – P. 171–182.
17. Nalda N.M.J., Yague B.J.L., Calva D.J.C., Gómez M.M.T. Classifying honeys from the Soria Province of Spain via multivariate analysis // *Anal. Bioanal. Chem.* – 2005. – 382. – P. 311–319.
18. Jurado J.M., Ballesteros O., Alcazar A., Pablos F., Martin M.J., Vilchez J.L., Navalon A. Differentiation of certified brands of origins of Spanish white wines by HS-SPME-GC and chemometrics // *Anal. Bioanal. Chem.* – 2008. – 390. – P. 961–970.
19. Cuny M., Vigneau E., Le Gall G., Colquhoun I., Less M., Rutledge D.N. Fruit juice authentication by ^1H NMR spectroscopy in combination with different chemometrics tools // *Anal. Bioanal. Chem.* – 2008. – 390. – P. 419–427.
20. Sumar S., Izmail H. Adulteration of foods – past and present // *Nutrition & Food Sci.* – 1995. – 95, № 4. – P. 11–15.
21. Arvanitoyannis I.S., Vaitsi O.B. A review on tomato authenticity: quality control methods in conjunction with multivariate analysis // *Crit. Rev. in Food Sci. and Nutrition* – 2007. – 47. – P. 675–699.
22. Рыков С.В., Корягина И.С., Скаковский Е.Д. Принципы контроля экологической и санитарно-гигиенической безопасности плодовых и ягодных соков // *Экологические нормы. Правила. Информация.* – 2009. – № 4. – С. 34–39.
23. Knödler M., Most M., Schieber A., Carle R. A novel approach to authenticity control of whole grain durum wheat (*Triticum durum* Desf.) flour and pasta, based on analysis of alkylresorcinol composition // *Food Chem.* – 2010. – 118, № 1. – P. 177–181.
24. Socaciu C., Ranga F., Fetea F., Leopold L., Dulf F., Parlog R. Complementary advanced techniques applied for plant and food authentication // *Czech J. of Food Sci.* – 2009. – 27. – P. 70–75.
25. Carter R.M., Yan Y., Tomlins K. Digital imaging based classification and authentication of granular food products // *Management Sci. and Technol.* – 2006. – 17, № 2. – P. 235–240.
26. Diaz-Maroto M.C., Castillo N., Castro-Vazquez L., De Torres C., Perez-Coello M.S. Authenticity evaluation of different mints based on their volatile composition and olfactory profile // *J. of Essential Oil-Bearing Plants.* – 2008. – 11, № 1. – P. 1–16.
27. Winterhalter P. Authentication of food and wine // *ACS Symposium Series.* – 2007. – 952. – P. 2–12.
28. Martin M., Martin G. Frauds in food products: A challenge for analytical chemistry // *Actualite Chimique.* – 2000. – 11. – P. 18–20.
29. Destailats F., de Wispelaere M., Joffre F., Golay P.-A., Hug B., Giuffrida F., Fauconnot L., Dionisi F. Authenticity of milk fat by fast analysis of triacylglycerols. Application to the detection of partially hydrogenated vegetable oils // *J. of Chromatography A.* – 2006. – 1131, № 1–2. – P. 227–234.
30. Fugel R., Carle R., Schieber A. Quality and authenticity control of fruit purees, fruit preparations and jams – A review // *Trends in Food Sci. and Technol.* – 2005. – 16, № 10. – P. 433–441.
31. Jee M. Oils and fats – authenticity and adulteration // *Food Sci. and Technol.* – 2004. – 18, № 1. – P. 28–29.
32. Zhang Y., Krueger D., Durst R., Lee R., Wang D., Seeram N., Heber D. International multidimensional authenticity specification (IMAS) algorithm for detection of commercial pomegranate juice adulteration // *J. of Agricult. and Food Chem.* – 2009. – 57, № 6. – P. 2550–2557.
33. Cotte J.F., Casabianca H., Chardon S., Lheritier J., Grenier-Loustalot M.F. Application of carbohydrate analysis to verify honey authenticity // *J. of Chromatography A.* – 2003. – 1021, № 1–2. – P. 145–155.
34. Dennis M.J. Establishing food authenticity // *Chem. and Ind. (London).* – 1997. – 24. – P. 997–1000.
35. Kelly J.F.D., Downey G., Fouratier V. Initial study of honey adulteration by sugar solutions using midinfrared (MIR) spectroscopy and chemometrics // *J. of Agricult. and Food Chem.* – 2004. – 52, № 1. – P. 33–39.
36. Marini F., Balestrieri F., Bucci R., Magri A.L., Marini D. Supervised pattern recognition to discriminate the geographical origin of rice bran oils: a first study // *Microchem. J.* – 2003. – 74. – P. 239–248.
37. Laursen K.H., Hansen T.H., Persson D.P., Schjoerring J.K., Husted S. Multi-elemental fingerprinting of plant tissue by semi-quantitative ICP-MS and Chemometrics // *J. of Anal. At. Spectrom.* – 2009. – 24, № 9. – P. 1198–1207.
38. Arvanitoyannis I.S., Katsota M.N., Psarra E.P., Soufleros E.H., Kallithraka S. Application of quality control methods for assessing wine authenticity: use of multivariate analysis (chemometrics) // *Trends in Food Sci. and Technol.* – 10, № 10. – P. 321–336.
39. Arvanitoyannis I.S., Tsitsika E.V., Panagiotaki P. Implementation of quality control methods (physicochemical, microbiological and sensory) in conjunction with multivariate analysis towards fish authenticity // *Int. J. of Food Sci. and Technol.* – 2005. – 40, № 3. – P. 237–263.
40. Marini F. Artificial neural networks in food analysis: trends and perspectives. A review // *Anal. Chim. Acta.* – 2009. – 635. – P. 121–131.
41. Marini F., Bucci R., Magri A.D., Magri A.L. Authentication of Italian CDO wines by class-

modeling techniques // *Chemometr. Intell. Lab.* – 2006. – 84. – P. 164–171.

42. Yashin Ya.I., Yashin A.Ya. Analysis of food products and beverages using high-performance liquid chromatography and ion chromatography with electrochemical detectors // *J. of Anal. Chem.* – 2004. – 59, № 12. – P. 1121–1127.

43. Savchuk S.A., Kolesov G.M., Nuzhnyi V.P. Chromatographic study of the chemical composition and potential toxicity of spirits and alcoholic beverages // *J. of Anal. Chem.* – 2007. – 62, № 12. – P. 641–649.

44. Savchuk S.A., Vlasov V.N., Appolonova S.A., Arbuzov V.N., Vedenin A.N., Mezinov A.B., Grigor'yan B.R. Application of chromatography and spectrometry to the authentication of alcoholic beverages. – *J. of Anal. Chem.* – 2001. – 56, № 3. – P. 214–231.

45. Arbuzov V.N., Savchuk S.A. Identification of vodkas by ion chromatography and gas chromatography // *J. of Anal. Chem.* – 2002. – 57, № 5. – P. 428–433.

46. Rudakov O.B., Polyanskii K.K., Alekseyuk M.P. Qualitative identification of milk fat from chromatographic data // *J. of Anal. Chem.* – 2002. – 57, № 12. – P. 1081–1088.

47. Savchuk S.A., Kolesov G.M. Chromatographic determination of phthalic acid esters as an indicator of adulterated cognacs and cognac spirits // *J. of Anal. Chem.* – 2007. – 62, № 8. – P. 761–772.

48. Kuchmenko T.A., Lisitskaya R.P., Bobrova O.S. Test method for detecting adulterated apple juice using piezoelectric resonators // *J. of Anal. Chem.* – 2008. – 63, № 6. – P. 596–602.

49. Karoui R., De Baerdemaeker J.A. A review of the analytical methods coupled with chemometric tools for the determination of the quality and identity of dairy products // *Food Chem.* – 2007. – 102. – P. 621–640.

50. De Leon-Rodríguez A., Escalante-Minakata P., Jiménez-García M.I., Ordonez-Acevedo L.G., Flores J.L.F., De La Rosa A.P.B. Characterization of volatile compounds from ethnic Agave alcoholic beverages by gas chromatography-mass spectrometry // *Food Technol. and Biotechnol.* – 2008. – 46, № 4. – P. 448–455.

51. Cabanero A.I., Recio J.L., Ruperez M. Liquid chromatography coupled to isotope ratio mass spectrometry: a new perspective on honey adulteration detection // *J. of Agricult. and Food Chem.* – 2006. – 54, № 26. – P. 9719–9727.

52. Calderone G., Guillou C., Reniero F., Naulet N. Helping to authenticate sparkling drinks with $^{13}\text{C}/^{12}\text{C}$ of CO_2 by gas chromatography-isotope ratio mass spectrometry // *Food Research Int.* – 2007. – 40, № 3. – P. 324–331.

53. Gomez-Ariza J.L., Villegas-Portero M.J., Bernal-Daza V. Characterization and analysis of amino acids in orange juice by HPLC-MS/MS for authenticity assessment // *Anal. Chim. Acta.* – 2005. – 540, № 1. – P. 221–230.

54. Nhu-Trang T.-T., Casabianca H., Grenier-Loustalot M.-F. Authenticity control of essential oils containing citronellal and citral by chiral and stable-isotope gas-chromatographic analysis // *Anal. and Bioanal. Chem.* – 2006. – 386, № 7–8. – P. 2141–2152.

55. Alonso-Salces R.M., Ndjoko K., Queiroz E.F., Ioset J.R., Hostettmann K., Berrueta L.A., Gallo B., Vicente F. On-line characterisation of apple polyphenols by liquid chromatography coupled with mass spectrometry and ultraviolet absorbance detection // *J. of Chromatography A.* – 2004. – 1046, № 1–2. – P. 89–100.

56. Csehati T., Forgacs E., Deyl Z., Miksik I. Chromatography in authenticity and traceability tests of vegetable oils and dairy products: A review // *Biomed. Chromatography.* – 2005. – 19, № 3. – P. 183–190.

57. Cotte J.F., Casabianca H., Giroud B., Albert M., Lheritier J., Grenier-Loustalot M.F. Characterization of honey amino acid profiles using high-pressure liquid chromatography to control authenticity // *Anal. and Bioanal. Chem.* – 2004. – 378, № 5. – P. 1342–1350.

58. von Baer D., Rentzsch M., Hitschfeld M.A., Mardones C., Vergara C., Winterhalter P. Relevance of chromatographic efficiency in varietal authenticity verification of red wines based on their anthocyanin profiles: interference of pyranoanthocyanins formed during wine ageing // *Anal. Chim. Acta.* – 2008. – 621, № 1. – P. 52–56.

59. Lachenmeier D.W., Richling E., Lopez M.G., Frank W., Schreier P. Multivariate analysis of FTIR and ion chromatographic data for the quality control of tequila // *J. of Agricult. and Food Chem.* – 2005. – 53, № 6. – P. 2151–2157.

60. Mosandl A. Capillary gas chromatography in quality assessment of flavours and fragrances // *J. of Chromatography.* – 1992. – 624, № 1–2. – P. 267–292.

61. Cunha S.C., Oliveira M.B.P.P. Discrimination of vegetable oils by triacylglycerols evaluation of profile using HPLC/ELSD // *Food Chem.* – 2006. – 95, № 3. – P. 518–524.

62. Kvasnicka F. Capillary electrophoresis in food authenticity // *J. of Separation Sci.* – 2005. – 28, № 9–10. – P. 813–825.

63. Golay P.-A., Giuffrida F., Dionisi F., Destailats F. Streamlined methods for the resolution and quantification of fatty acids including trans fatty acid

isomers in food products by gas chromatography // *J. of AOAC Int.* – 2009. – 92, № 5. – P. 1301–1309.

64. Scharrer A., Mosandl A. Reinvestigation of vanillin contents and component ratios of vanilla extracts using high-performance liquid chromatography and gas chromatography // *Deutsche Lebensmittel-Rundschau.* – 2001. – 97, № 12. – P. 449–456.

65. Schieber A., Keller P., Carle R. Determination of phenolic acids and flavonoids of apple and pear by high-performance liquid chromatography // *J. of Chromatography A.* – 2001. – 910, № 2. – P. 265–273.

66. Low N.H., McLaughlin M., Hofsommer H.-J., Hammond D.A. Capillary gas chromatographic detection of invert sugar in heated, adulterated, and adulterated and heated apple juice concentrates employing the equilibrium method // *J. of Agricult. and Food Chem.* – 1999. – 47, № 10. – P. 4261–4266.

67. Liang L., Liu Z.-X., Yang M.-H., Zhang Y.-X., Wang C.-H. Discrimination of variety and authenticity for rice based on visual/near infrared reflection spectra // *J. of Infrared and Millimeter Waves.* – 2009. – 28, № 5. – P. 353–356.

68. Chen L.-Z., Zhao J., Ye Z.-H., Zhong Y.-P. Determination of adulteration in honey using near-infrared spectroscopy // *Spectrosc. and Spectral Analysis.* – 2008. – 28, № 11. – P. 2565–2568.

69. Nicolaou N., Goodacre R. Rapid and quantitative detection of the microbial spoilage in milk using Fourier transform infrared spectroscopy and chemometrics // *Analyst.* – 2008. – 133. – P. 1424–1431.

70. Woo Y.-Ah, Kim H.-J., Chung H. Classification of cultivation area of ginseng radix with NIR and Raman spectroscopy // *Analyst.* – 1999. – 124. – P. 1223–1226.

71. Boffo E.F., Tavares L.A., Ferreira M.M.C., Ferreira A.G. Classification of Brazilian vinegars according to their ¹H NMR spectra by pattern recognition analysis // *LWT – Food Sci. and Technol.* – 2009. – 42, № 9. – P. 1455–1460.

72. Viggiani L., Morelli M.A.C. Characterization of wines by nuclear magnetic resonance: a work study on wines from the Basilicata region in Italy // *J. of Agricult. and Food Chem.* – 2008. – 56, № 18. – P. 8273–8279.

73. Casale M., Armanino C., Casolino C., Forina M. Combining information from headspace mass spectrometry and visible spectroscopy in the classification of the Ligurian olive oils // *Anal. Chim. Acta.* – 2007. – 589, № 1. – P. 89–95.

74. Shintu L., Caldarelli S., Franke B.M. Preselection of potential molecular markers for the geographic origin of dried beef by HR-MAS NMR spectroscopy // *Meat Sci.* – 2007. – 76, № 4. – P. 700–707.

75. Cabañero A.I., Recio J.L., Rupérez M. Isotope ratio mass spectrometry coupled to liquid and gas chromatography for wine ethanol characterization // *Rapid Commun. in Mass Spectrom.* – 2008. – 22, № 20. – P. 3111–3118.

76. Greule M., Hansel C., Bauermann U., Mosandl A. Feed additives: Authenticity assessment using multicomponent-/multielement- isotope ratio mass spectrometry // *European Food Research and Technol.* – 2008. – 227, № 3. – P. 767–776.

77. Richling E., Appel M., Heckel F., Kahle K., Kraus M., Preston C., Hummer W., Schreier P. Flavor authenticity studies by isotope ratio mass spectrometry: perspectives and limits // *ACS Symposium Series.* – 2007. – 952. – P. 75–86.

78. Zeynali F., Tajik H., Asri-Rezaie S., Meshkini S., Fallah A.A., Rahnama M. Determination of copper, zinc and iron levels in edible muscle of three commercial fish species from Iranian coastal waters of the Caspian sea // *J. of Animal and Veterinary Advances.* – 2009. – 8, № 7. – P. 1285–1288.

79. Khan S., Kazi T.G., Baig J.A., Kolachi N.F., Afridi H.I., Shah A.Q., Kandhro G.A., Kumar S. Separation and preconcentration of trace amounts of aluminum ions in surface water samples using different analytical techniques // *Talanta.* – 2009. – 80, № 1. – P. 158–162.

80. Rizzon L.A., Miele A. Analytical characteristics of Merlot wines from the Serra Gaúcha region [Características analíticas de vinhos Merlot da Serra Gaúcha] // *Ciencia Rural.* – 2009. – 39, № 6. – P. 1913–1916.

81. Scott S.M., James D., Ali Z., O'Hare W.T., Rowell F.J. Total luminescence spectroscopy with pattern recognition for classification of edible oils // *Analyst.* – 2003. – 128. – P. 966–973.

82. Cuadrado U.M., de Castro L.M.D., Gomez-Nieto M.A. Study of spectral analytical data using fingerprints and scaled similarity measurements // *Anal. Bioanal. Chem.* – 2005. – 381. – P. 953–963.

83. Woodcock T., Downey G., O'Donnell C.P. Near infrared spectral fingerprinting for confirmation of claimed PDO provenance of honey // *Food Chem.* – 2009. – 114, № 2. – P. 742–746.

84. Vlasov Y., Legin A. Non-selective chemical sensors in analytical chemistry: from "electronic nose" to "electronic tongue" // *Fresenius J. Anal. Chem.* – 1998. – 361. – P. 255–260.

85. Mielle P., Tarrega A., Gorria P., Liodenot J.J., Liaboef J., Andrejewski J.-L., Salles C. From human to artificial mouth, from basics to results // *AIP Conference Proceedings.* – 2009. – 1137. – P. 144–147.

86. Leake L.L. Electronic noses and tongues // *Food Technol.* – 2006. – 60, № 6. – P. 96–102.

87. Lvova L., Kim S.S., Legin A., Vlasov Y., Yang J.S., Cha G.S., Nam H. All-solid-state electronic

- tongue and its application for beverage analysis // *Anal. Chim. Acta.* – 2002. – 468. – P. 303–314.
88. Winquist F., Wide P., Lundström I. An electronic tongue based on voltammetry // *Anal. Chim. Acta.* – 1997. – 357. – P. 21–31.
89. Winquist F., Krantz-Rulcker C., Lundstrom I. Electronic tongues // *MRS Bulletin.* – 2004. – 29, № 10. – P. 726–731.
90. Ciosek P., Brzozka Z., Wroblewski W. Classification of beverages using a reduced sensor array // *Sensors and Actuators B: Chemical.* – 2004. – 103, № 1–2. – P. 76–83.
91. Buratti S., Benedetti S., Scampicchio M., Pangerod E.C. Characterization and classification of Italian Barbera wines by using an electronic nose and an amperometric electronic tongue // *Anal. Chim. Acta.* – 2004. – 525, № 1. – P. 133–139.
92. Tan T., Schmitt V., Isz S. Electronic tongue: a new dimension in sensory analysis: instrument complements use of the electronic nose for sensory analysis of food flavor by measuring nonvolatile flavor components // *Food Technol.* – 2001. – 55, № 10. – P. 44–50.
93. Bleibaum R.N., Stone H., Tan T., Labreche S., Saint-Martin E., Isz S. Comparison of sensory and consumer results with electronic nose and tongue sensors for apple juices // *Food Quality and Preference.* – 2002. – 13, № 6. – P. 409–422.
94. Родионова О.Е., Померанцев А.Л. Хемометрика в аналитической химии, 2006. – 61 с. http://www.chemometrics.ru/materials/articles/chemometrics_review.pdf
95. Дронов С.В. Многомерный статистический анализ: Учебное пособие. – Барнаул: Изд-во Алт. гос. ун-та, 2003. – С. 88–89.
96. Померанцев А.Л. Метод главных компонент(PCA), 2008. <http://www.chemometrics.ru/materials/textbooks/pca.htm>
97. Antti H. Multivariate characterization of wood related materials. PhD Thesis. Umea University, 1999. – 86 p.
98. Hopke P.K. The evolution of chemometrics // *Anal. Chim. Acta.* – 2003. – 500. – P. 365–377.
99. Legin A., Rudnitskaya A., Clapham D., Seleznev B., Vlasov Y. Electronic tongue for pharmaceutical analytics: quantification of tastes and masking effects // *Anal. Bioanal. Chem.* – 2004. – 380. – P. 36–45.
100. Hruskar M., Major N., Krpan M., Krbavcic I.P., Šarić G., Marković K., Vahčić N. Evaluation of milk and dairy products by electronic tongue // *Mljekarstvo.* – 2009. – 59, № 3. – P. 193–200.
101. Oliveri P., Baldo M.A., Daniele S., Forina M. Development of a voltammetric electronic tongue for discrimination of edible oils // *Anal. Bioanal. Chem.* – 2009. – 395. – P. 1135–1143.
102. Cotte J.F., Casabianca H., Chardon S., Lheritier J., Grenier-Loustalot M.F. Chromatographic analysis of sugars applied to the characterization of monofloral honey // *Anal. Bioanal. Chem.* – 2004. – 380. – P. 698–705.
103. Li H., Takahashi Y., Kumagai M., Fujiwara K., Kikuchi R., Yoshimura N., Amano T., Ogawa N. A chemometrics approach for distinguishing between beers using near infrared spectroscopy // *J. of Near Infrared Spectrosc.* – 2009. – 17, № 2. – P. 69–76.
104. Shin Y.-S., Bang K.-H., In D.-S., Kim O.-T., Hyun D.-Y., Ahn I.-O., Bon C.K., Choi H.-K. Fingerprinting analysis of fresh ginseng roots of different ages using ¹H-NMR spectroscopy and principal components analysis // *Archives of Pharmacal Research.* – 2007. – 30, № 12. – P. 1625–1628.
105. Urmano M., de Castro M.D.L., Perez P.M., Garcia-Olmo J., Gomez-Nieto M.A. Ultraviolet-visible spectroscopy and pattern recognition methods for differentiation and classification of wines // *Food Chem.* – 2006. – 97. – P. 166–175.
106. Cozzolino D., Vadell A., Ballesteros F., Galletta G., Barlocco N. Combining visible and near-infrared spectroscopy with Chemometrics to trace muscles from an autochthonous breed of pig produced in Uruguay: a feasibility study // *Anal. Bioanal. Chem.* – 2006. – 385. – P. 931–936.
107. Родионова О.Е. Задачи классификации и дискриминации. Презентация лекции. Омск, Омский гос. ун-т им. Ф.М. Достоевского, 2007. <http://www.chemometrics.ru/materials/presentations/omsk-2007/Classification.ppt>
108. Кучерявский С.В. Методы многомерной классификации. Презентация лекции. Пятый межд. зимний симпозиум по хемометрике "Современные методы анализа многомерных данных", Самара, 2006. <http://www.chemometrics.ru/wsc5/docs/lecture03.ppt>
109. Родионова О.Е. Хемометрический подход к исследованию больших массивов химических данных // *Рос. хим. журн. (Журн. Рос. хим. об-ва им. Д.И. Менделеева).* – 2006. – 4, № 2. – С. 128–144.
110. Stanimirova I., Ustun B., Cajka T., Riddelova K., Hajslova J., Buydens L.M.C., Walczak B. Tracing the geographical origin of honeys based on volatile compounds profiles assessment using pattern recognition techniques // *Food Chem.* – 2010. – 118, № 1. – P. 171–176.
111. Shiroma C., Rodriguez-Saona L. Application of NIR and MIR spectroscopy in quality control of potato chips // *J. of Food Composition and Analysis.* – 2009. – 22, № 6. – P. 596–605.
112. Jaganathan J., Mabud Md. A., Dugar S. Geographic origin of wine via trace and ultra-trace elemental analysis using inductively coupled plasma

mass spectrometry and chemometrics // ACS Symposium Series. – 2007. – 952. – P. 200–206.

113. Xie L.-J., Ying Y.-B. Use of near-infrared spectroscopy and least-squares support vector machine to determine quality change of tomato juice // J. of Zhejiang University: Science B. – 2009. – 10, № 6. – P. 465–471.

114. Woodcock T., Downey G., Kelly J.D., O'Donnell C. Geographical classification of honey samples by near-infrared spectroscopy: a feasibility study // J. of Agricult. and Food Chem. – 2007. – 55, № 22. – P. 9128–9134.

115. Chen Q., Zhao J., Zhang H., Wang X. Feasibility study on qualitative and quantitative analysis in tea by near infrared spectroscopy with multivariate calibration // Anal. Chim. Acta. – 2006. – 572, № 1. – P. 77–84.

116. Tian S.-Y., Deng S.-P., Ding C.-H., Yin C.-L., Li H. Discrimination of red wine age using voltammetric electronic tongue based on multifrequency large-amplitude voltammetry and pattern recognition method // Sensors and Materials. – 2007. – 19, № 5. – P. 287–298.

117. He J., Rodriguez-Saona L.E., Giusti M.M. Midinfrared spectroscopy for juice authentication – rapid differentiation of commercial juices // J. of Agricult. and Food Chem. – 2007. – 55, № 11. – P. 4443–4452.

118. Chen Z.-P., Jiang J.-H., Li Y., Liang Yi-Z., Ru-Qin Yu. Fuzzy linear discriminant analysis for chemical data sets // Chemometr. and Intell. Lab. Systems. – 1999. – 45. – P. 295–302.

119. Факторный, дискриминантный и кластерный анализ: Пер. с англ. / Ким Дж.-О., Мьюллер Ч.У., Клекка У.Р. и др.; Под ред. И.С. Енюкова. – М.: Финансы и статистика, 1989. – 215 с.

120. Прикладная статистика: Классификация и снижение размерности: / Айвазян С.А., Бухштабер В.М., Енюков И.С., Мешалкин Л.Д. – М.: Финансы и статистика, 1989. – 607 с.

121. Кендалл М. Многомерный статистический анализ и временные ряды: / Кендалл М., Стюарт А., Пресман Э.Л., Ротарь В.И., Колмогоров А.Н., Прохоров Ю.В. – М.: Наука, 1976. – 736 с.

122. Дубров А.М. Многомерные статистические методы. / Дубров А.М., Мхитарян В.С., Трошин Л.И. – М.: Финансы и статистика, 2000. – 350 с.

123. Афифи А. Статистический анализ: Подход с использованием ЭВМ: Пер. с англ. / Афифи А., Эйзен С., Енюков И.С., Новиков И.Д., Башарин Г.П. – М.: Мир, 1982. – 488 с.

124. Линейный регрессионный анализ результатов химического эксперимента в системе Statistica. Учебно-методическое пособие / Михальчук В.М., Михальчук А.В. / Донецк: ДонНУ, 2002. –

66с. http://www.donnu.edu.ua/chem/student/methodic/line_analysis.pdf

125. Налимов В.В. Применение математической статистики при анализе вещества. – М.: Гос. изд-во физ.-мат. л-ры, 1960. – 430 с.

126. Paradkar M.M., Irudayaraj J., Sakhamuri S. Discrimination and classification of beet and cane sugars and their inverts in maple syrup by FT-Raman spectroscopy // Applied Eng. in Agricult. – 2002. – 18, № 3. – P. 379–383.

127. Suchánek M., Filipová H., Volka K., Delgado I., Davies A.N. Qualitative analysis of green coffee by infrared spectrometry // Fresenius J. of Anal. Chem. – 1996. – 354, № 3. – P. 327–332.

128. Cozzolino D., Parker M., Damberg R.G., Herderich M., Gishen M. Chemometrics and visible-near infrared spectroscopic monitoring of red wine fermentation in a pilot scale // Biotechnol. and Bioeng. – 2006. – 95, № 6. – P. 1101–1107.

129. Granitto P.M., Gasperi F., Biasioli F., Trainotti E., Furlanello C. Modern data mining tools in descriptive sensory analysis: a case study a Random forest approach // Food Quality and Preference. – 2007. – 78. – P. 681–689.

130. Timm H., Borgelt C., Döring C., Kruse R. An extension to possibilistic fuzzy cluster analysis // Fuzzy Sets and Syst. – 2004. – 147. – P. 3–16.

131. Tsai C.-An, Lee Te-C., Ho I.-C., Yang U.-C., Chen C.-H., Chen J.J. Multi-class clustering and prediction in the analysis of microarray data // Mathemat. Biosci. – 2005. – 193. – P. 79–100.

132. Leski J. Towards a robust clustering // Fuzzy Sets and Syst. – 2003. – 137. – P. 215–233.

133. Dumitrescu D., Lazzerini B., Jain L.C. Fuzzy sets and their application to clustering and training. – CRC Press LLC: Florida, 2000. – 593 с.

134. Zahid N., Abouela O., Limouri M., Essaid A. Fuzzy clustering on k-nearest-neighbours rule // Fuzzy Sets and Syst. – 2001. – 120. – P. 239–247.

135. Wu K.-L., Yang M.-S. Alternative c-means clustering algorithms // Pattern Recognition. – 2002. – 35. – P. 2267–2278.

136. Shahin M.A., Verma B.P., Tollner E.W. Fuzzy logic model for predicting peanut maturity // Transactions of the American Soc. of Agricult. Engineers. – 2000. – 43, № 2. – P. 483–490.

137. Андреев И.М. Описание алгоритма CART // Методы. Алгоритмы. Программы. – 2004. – 3–4, № 7–8. – С. 48–53.

138. Шитиков В.К., Розенберг Г.С., Зинченко Т.Д. Количественная гидроэкология: методы системной идентификации. – Тольятти: Институт экологии Волжского бассейна РАН, 2003. – 463 с. <http://www.ievbran.ru/kiril/Library/Book1/content0/content0.htm>

139. Zhang M. H., Xu Q.S., Daeyaert F., Lewi P.J., Massart D.L. Application of boosting to classification problems in chemometrics // *Anal. Chim. Acta.* – 2005. – 544. – P. 167–176.
140. Dobra A.V. Classification and regression tree construction. PhD Thesis. – Cornell University, 2003.–186p.
<http://www.cise.ufl.edu/~adobra/old/papers/phd-thesis.pdf>
141. Шампандер А. Дж. Искусственный интеллект в компьютерных играх: как обучить виртуальные персонажи реагировать на внешние воздействия. Гл. 26 Деревья классификации и регрессии. – М.: Вильямс, 2007. – 385–401 с.
<http://www.williamspublishing.com/PDF/978-5-8459-1170-4/part.pdf>
142. Khoshgoftaar T.M., Allen E.B., Naik A., Jones W.D., Hudepohl J.P. Using classification trees for software quality models: Lessons learned // *Int. J. of Software Eng. and Knowledge Eng.* – 1999. – 9, № 2. – P. 217–231.
143. Smeti E.M., Thanasoulis N.C., Lytras E.S., Tzoumerkas P.C., Golfopoulos S.K. Treated water quality assurance and description of distribution networks by multivariate chemometrics // *Water Research.* – 2009. – 43, № 18. – P. 4676–4684.
144. Ortiz M.C., Herrero A., Sanchez M.S., Saraibia L.A., Iniguez M. Modelling the relation between CieLab parameters and sensory scores for quality control of red-wine colour // *Analyst.* – 1995. – 120, № 12. – P. 2793–2798.
145. Kokkinofa R.I., Theocharis C.R. Chemometric characterization of the cypriot spirit "Zivania" // *J. of Agricult. and Food Chem.* – 2005. – 53, № 13. – P. 5067–5073.
146. Mutihac L., Mutihac R. Mining in chemometrics // *Anal. Chim. Acta.* – 2008. – 612. – P. 1–18.
147. Nobel W.S. What is a support vector machine? // *Nature Biotechnol.* – 2006. – 24, № 12. – P. 1565–1567.
148. Breton R.G., Lloyd G.R. Support Vector Machines for classification and regression // *Analyst.* – 2010. – 135. – P. 230–267.
149. Gunn S. Support vector machines for classification and regression. ISIS Technical Report. – University of Southampton, 1997. – 42 p.
<http://www.svms.org/tutorials/Gunn1997.pdf>
150. Asano A. Support vector machine and kernel method // *Pattern information processing (2004 Autumn Semester) Session 12.* – 3 p.
<http://www.svms.org/tutorials/Asano2004.pdf>
151. Лившиц Ю. Метод опорных векторов: лекция № 7 курса "Алгоритмы для Интернета", 2006. – 9 с. <http://yury.name/internet/07ianote.pdf>
152. Lin Y. On the support vector machines. Technical Report № 1029. – Madison: University of Wisconsin, Department of Statistics, 2000. – 25 p.
<http://www.svms.org/tutorials/Lin2000.pdf>
153. Воронцов К.В. Лекции по методу опорных векторов, 2007. – 18 с.
<http://www.ccas.ru/voron/download/SVM.pdf>
154. Браславский П., Густелев В. Система автоматического реферирования новостных сообщений на основе машинного обучения. // Электронные библиотеки: перспективные методы и технологии, электронные коллекции. Труды 9-й Всеросс. конф. RCDL'2007. – Переславль-Залесский: Изд-во "Университет города Переславля", 2007.– С. 142–147.
<http://www.kansas.ru/pb/paper/rcdl2007.pdf>
155. Erästö P. Support vector machines – backgrounds and practice. Dissertation for the Degree of Licentiate of Philosophy. – Helsinki: Rolf Nevanlinna Institute, 2001. – 78 p.
<http://www.svms.org/tutorials/Erasto2001.pdf>
156. Burges C.J.C. A tutorial on support vector machines for pattern recognition // *Data Mining and Knowledge Discovery.* – 1998. – 2, № 2. – P. 121–167.
157. Burbidge R. Support vector machines: some perspectives from probability and statistics // *Imperial College Statistics Seminars*, 2001. – 53 p.
<http://www.svms.org/tutorials/Burbidge2001.pdf>
158. El Barbri N., Llobet E., El Bari N., Correig X., Bouchikhi B. Electronic nose based on metal oxide semiconductor sensors as an alternative technique for the spoilage classification of red meat // *Sensors.* – 2008. – 8, № 1. – P. 142–156.
159. Unay D., Gosselin B. Stem and calyx recognition on Jonagold apples by pattern recognition // *J. of Food Eng.* – 2007. – 78, № 2. – P. 597–605.
160. Yang H., Zhou Z., Zhao R., Wang B. Recognition of wheat varieties by image analysis // *Proceedings of SPIE – The Int. Society for Optical Eng.* – 2003. – 5286, № 1. – P. 465–468.
161. Setiawan M.A., Rao R.K., Lakshminarayanan S. Partial correlation metric based classifier for food product characterization // *J. of Food Eng.* – 2009. – 90. – P. 146–152.
162. Эсбенсен К. Анализ многомерных данных. Избранные главы / Пер. с англ. С.В. Кучерявского под ред. О.Е. Родионовой. – Черногловка: Институт проблем химической физики РАН, 2005. – 160 с.
163. Родионова О., Померанцев А. Проекционные методы в системе Excel: Учебное пособие // Российское хемометрическое общество.
<http://www.chemometrics.ru/materials/textbooks/projection.htm#Ch1.3>
164. Lakshminarayanan S., Sirish L., Nandakumar K. Modelling and control of multivariable process:

the dynamic projection to latent structures approach // *AIChE J.* – 1997. – 43. – P. 2307–2323.

165. Abdi, H. Partial least square regression, projection on latent structure regression (PLS-Regression) // *Wiley Interdisciplinary Reviews: Computational Statistics*, V. 2, Iss. – Wiley, 2010. – P. 97–106. <http://www.utdallas.edu/~herve/abdi-wireCS-PLS2010.pdf>

166. Nunes A., Martins J., Barros A.S., Galvis-Sánchez A.C., Delgadillo I. Estimation of olive oil acidity using FT-IR and partial least squares regression // *Sensing and Instrumentation for Food Quality and Safety*. – 2009. – 3, № 3. – P. 187–191.

167. Lachenmeier D.W. Rapid quality control of spirit drinks and beer using multivariate data analysis of Fourier transform infrared spectra // *Food Chem.* – 2007. – 101, № 2. – P. 825–832.

168. Lin P., Chen Y.M., He Y. Fast discrimination of varieties of sugar based on spectroscopy technology // *Spectrosc. and Spectral Analysis*. – 2009. – 29, № 2. – P. 382–385.

169. Armenta S., Garrigues S., de la Guardia M. Determination of edible oil parameters by near infrared spectrometry // *Anal. Chim. Acta.* – 2007. – 596, № 2. – P. 330–337.

170. Mizrach A., Schmilovitch Z., Korotic R., Iru-dayaraj J., Shapira R. Yeast detection in apple juice using Raman spectroscopy and chemometric methods // *Transactions of the ASABE*. – 2007. – 50, № 6. – P. 2143–2149.

171. Fagan C.C., O'Donnell C.P., O'Callaghan D.J., Downey G., Sheehan E.M., Delahunty C.M., Everard C., Guinee T.P., Howard V. Application of mid-infrared spectroscopy to the prediction of maturity and sensory texture attributes of cheddar cheese // *J. of Food Sci.* – 2007. – 72, № 3. – P. E130–E137.

172. Andrade J.M., Gomez-Carracedo M.P., Fernandez E., Elbergali A., Kubista M., Prada D. Classification of commercial apple beverages using a minimum set of mid-IR wavenumbers selected by Procrustes rotation // *Analyst*. – 2003. – 128. – P. 1193–1199.

173. Ruth S., Kiers J.L., Akkermans W., Perez R., Koot A.H., Perri E., Pellegrino M., Moreno Rojas J.M., Guillou C., Rossignol-Castera A. Geographical origin, cultivar and harvesting year verification of European and non-European olive oils using proton transfer reaction mass spectrometry with multivariate data analysis // *Proceedings of the 5th Intern. Technical Symposium on Food Processing, Monitoring Technol. in Bioprocesses and Food Quality Management*. – Potsdam, Germany. – 2009. – P. 791–797.

174. Ruth S., Koot A., Akkermans W., Araghipour N., Rozijn M., Baltussen M., Wisthaler A., Märk T., Frankhuizen R. Butter and butter oil classification by

PTR-MS // *European Food Research and Technol.* – 2008. – 227, № 1. – P. 307–317.

175. Galtier O., Dupuy N., Le Dréau Y., Ollivier D., Pinatel C., Kister J., Artaud J. Geographic origins and compositions of virgin olive oils determined by chemometric analysis of NIR spectra // *Anal. Chim. Acta.* – 2007. – 595, № 1–2. – P. 136–144.

176. Родионова О.Е. Интервальный метод обработки результатов многоканальных экспериментов. Дис. док. физ.-мат. наук: 01.04.01 / Ин-т хим. физики им. Н.Н. Семенова РАН. – М., 2008. – 272 с. www.nsc.ru/interval/Library/AppIDiss/Rodionova.pdf

177. Agatonovic-Kustrin S., Beresford R. Basic concepts of artificial neural network (ANN) modeling and its application in pharmaceutical research // *J. of Pharmaceutical and Biomed. Analysis*. – 2000. – 22. – P. 717–727.

178. Arbib M.A. The handbook of brain theory and neural networks, 2nd edn. – Cambridge, e.a.: Bradford Book, 2002. – 1290 p.

179. Gupta M.M., Jin L., Homma N. Static and dynamic neural network: from fundamentals to advanced theory. – Wiley, 2003. – 722 p.

180. Braspenning P.J., Thuijsman F., Weijters A.J.M.M. Artificial neural networks. An introduction to ANN theory and practice. – Berlin: Springer-Verlag, 1995. – 136 p.

181. Хайкин С. Нейронные сети: полный курс, 2-е изд., испр.: Пер. с англ. – М.: Вильямс, 2006. – 1104 с.

182. Круглов В.В., Борисов В.В. Искусственные нейронные сети. Теория и практика. 2-е изд., стереотип. – М.: Горячая линия – Телеком, 2002. – 382 с.

183. Осовский С. Нейронные сети для обработки информации / Пер. с польского И.Д. Рудинского. – М.: Финансы и статистика, 2002. – 344 с.

184. Freeman J.A. Neural networks: algorithms, applications and programming techniques. – New York, e.a.: Addison-Wesley, 1991. – 401 p.

185. Mandic D.P., Chambers J.A. Recurrent neural network for prediction: learning algorithms, architectures and stability (adaptive and learning systems for signal processing, communications and control series). – Wiley, 2001. – 285 p.

186. Krogh A. What are artificial neural networks? // *Nature Biotechnol.* – 2008. – 26, № 2. – P. 195–197.

187. Smits J.R.M., Melssen W.J., Buydens L.M.C., Kateman G. Using artificial neural networks for solving chemical problems. Part I. Multi-layer feed-forward networks // *Chemometr. and Intell. Lab. Systems*. – 1994. – 22, № 2. – P. 165–189.

188. Melssen W.J., Smits J.R.M., Buydens L.M.C., Kateman G. Using artificial neural networks for solving chemical problems. Part II. Kohonen self-

- organising feature maps and Hopfield networks // *Chemometr. and Intell. Lab. Systems.* – 1994. – 23, № 2. – P. 267–291.
189. Marini F. Artificial neural networks in foodstuff analyses: Trends and perspectives. A review // *Anal. Chim. Acta.* – 2009. – 635, № 2. – P. 121–131.
190. Llobet E., Hines E.L., Gardner J.W., Bartlett P.N., Mottram T.T. Fuzzy ARTMAP based electronic nose data analysis // *Sensors and Actuators B: Chemical.* – 1999. – 61, № 1. – P. 183–190.
191. Grošelj N., van der Veer G., Tušar M., Vračko M., Novič M. Verification of the geological origin of bottled mineral water using artificial neural networks // *Food Chem.* – 2010. – 118, № 4. – P. 941–947.
192. Wei B.-L., Chen G.-Q., Wang J., Xu J.-C., Yan G.-F., Zhu T., Gao S.-M. A new method for determining the concentration of melamine in milk // *Guangdianzi Jiguang / J. of Optoelectronics Laser.* – 2009. – 20, № 12. – P. 1618–1621.
193. da Costa Filho P.A. Rapid determination of sucrose in chocolate mass using near infrared spectroscopy // *Anal. Chim. Acta.* – 2009. – 631, № 2. – P. 206–211.
194. Francelin R.A., Gomide F.A.C., Lancas F.M. Use of artificial neural networks for the classification of vegetable oils after CG analysis // *Chromatography.* – 1993. – 35, № 3–4. – P. 160–166.
195. Cantero M.C., Perez R.M., Martinez P.J., Aguilar P.L., Plaza J., Plaza A. Analysis of the behaviour of a neural network model in the identification and quantification of hyperspectral signatures applied to the determination of water quality // *Proc. of SPIE. – The Int. Soc. for Optical Eng.* – 2004. – 5584, № 23. – P. 174–185.
196. Onea A., Collewet G., Fernandez C., Vertan C., Richard N., Mariette F. Quality analysis of blue-veined cheeses by MRI: a preliminary study // *Proc. SPIE. – The Int. Soc. for Optical Eng.* – 2003. – 5132. – P. 400–409.