

УДК 535.243+663.2

*Н. Ф. ДЕГТЯРЬ¹, Е. И. НЕЗАЛЬЗОВА¹, М. В. РОГОВАЯ², Г. В. СИНИЦЫН²,
Е. А. СКОРБАНОВА¹, М. А. ХОДАСЕВИЧ²*

**ИСПОЛЬЗОВАНИЕ МЕТОДОВ СПЕКТРОСКОПИИ УФ-,
ВИДИМОГО И БЛИЖНЕГО ИК-ДИАПАЗОНОВ ДЛЯ ОПРЕДЕЛЕНИЯ
ОРИГИНАЛЬНОСТИ ВЫДЕРЖАННЫХ ВИННЫХ ДИСТИЛЛЯТОВ**

*¹Научно-практический институт плодоводства, виноградарства и пищевых технологий
(Кишинев, Молдова)*

²Институт физики им. Б. И. Степанова НАН Беларуси

(Поступила в редакцию 11.07.2014)

В настоящее время все большее количество потребителей проявляют заботу о своем здоровье и хотят покупать натуральные и аутентичные продукты питания. Согласно [1], аутентичность (подлинность) пищевой продукции – неотъемлемая составная часть качества пищевой продукции, определяемая совокупностью физико-химических и биологических показателей, абсолютные количественные значения и интервалы изменения которых обоснованы природными свойствами сырья и допустимым технологическим воздействием при получении готовых пищевых продуктов. Из предложенной в работе [2] классификации видов подлинности нас в первую очередь интересует аутентичность оригинальности, которая подчеркивает исторические истоки происхождения товара или бренда, использование технологий или сырья, специфичных для данного вида продукции и отличных от общепринятой практики товарного производства.

Достаточно остро проблема аутентичности оригинальности проявляется в производстве и контроле качества коньяков и бренди, производимых из выдержанных винных дистиллятов. Основными факторами, препятствующими их фальсификации и выпуску низкокачественной продукции, являются контроль возраста выдержки и географического происхождения дистиллятов и идентификация производителя коньяков или бренди.

В данной статье в качестве объектов исследования были выбраны выдержанные молдавские винные дистилляты, характеризующиеся более однородными физико-химическими параметрами по сравнению с кондиционными по качеству дивинами (молдавскими бренди), в процессе купажа которых проводится смешивание дистиллятов разного возраста, воды, сахарного сиропа и красителя (колера).

Для определения аутентичности оригинальности исследуемых объектов нами применялась спектроскопия УФ-, видимого и ближнего ИК-диапазонов с последующей хемометрической обработкой спектров пропускания.

Именно спектроскопия ближнего ИК-диапазона в последнее время стала широко внедряться в фармацевтическую [3], нефтехимическую [4], пищевую [5, 6] отрасли промышленности, биомедицину [7, 8] и т. д. не только из-за развитости аппаратных средств, простоты и экспрессного характера их использования, но и из-за повышения качества получаемых результатов вследствие применения хемометрических методов. Наиболее удачно, с нашей точки зрения, понятие «хемометрика» определено в работе [9] – это дисциплина, использующая математические, статистические и другие методы на основе формальной логики для построения или выбора оптимальных процедур или экспериментов, а также для достижения максимума релевантной информации при анализе экспериментальных данных. Именно с помощью хемометрики спектроскопия ближнего ИК-диапазона, ранее мало востребованная вследствие зашумленности спектров из-за интенсивного

поглощения воды и влияния рассеяния света [10], стала наравне конкурировать со спектроскопией среднего ИК-диапазона для определения физических и химических характеристик исследуемых объектов.

Единственными общепринятыми оптическими характеристиками коньяков на сегодняшний день являются оптические плотности на длинах волн 420 и 520 нм [11]. Однако наличие в винных дистиллятах веществ, обладающих сходными оптическими свойствами, не позволяет сделать объективный вывод о качестве и свойствах исследуемых образцов по спектрофотометрическим измерениям на малом количестве выделенных длин волн. Поэтому в настоящей работе мы применим хемометрический подход к зарегистрированным двухлучевым спектрофотометром Shimadzu PC 3101 широкополосным спектрам пропускания (190–2600 нм, шаг сканирования 0,5 нм в диапазоне 190–480 нм и 1 нм в диапазоне 480–2600 нм) репрезентативной выборки из 42 молдавских винных дистиллятов и оценим качество моделирования характеристик исследуемых объектов. После регистрации спектры сглаживались фильтром Савицкого – Голея 3-й степени по 9 соседним отсчетам.

Таким образом, каждый исследуемый образец характеризуется 2698 спектральными отсчетами или переменными, т. е. он может быть представлен уникальной точкой в 2698-мерном пространстве. Для декомпозиции многомерного пространства нами применялся метод главных компонент (англ. principal component (PC) analysis, PCA) с предварительным центрированием данных (приведение средней величины отсчетов на каждой длине волны к нулевому значению) или автошкалированием (комбинация центрирования и нормирования на каждой длине волны на стандартное отклонение по выборке), которое уравнивает вклад разных переменных в рассматриваемую модель. PCA [12] последовательно выбирает направления PC в многомерном пространстве отсчетов, которым присуща максимальная дисперсия исходных данных. Для определения правильного числа PC был применен анализ зависимости объясненной дисперсии матрицы данных от числа PC. Установлено, что объясненная четырьмя главными компонентами дисперсия составляет 92,1 и 98,9 % для автошкалирования и центрирования соответственно.

Одним из признаков аутентичности коньяка или бренди является идентификация предприятия-производителя дистиллятов. Именно этот признак отражен в символах объектов на рис. 1. На нем не изображены дистилляты производителей, которые представлены в исследуемой выборке единственным образцом, и прослеживается неполная кластеризация исследуемых дистиллятов. Предварительный анализ этого рисунка вызывает сомнения в достоверной классификации обозначенных квадратами дистиллятов только одного производителя. Для более аргументированного вывода проведем классификацию исследуемых объектов по их спектральным данным

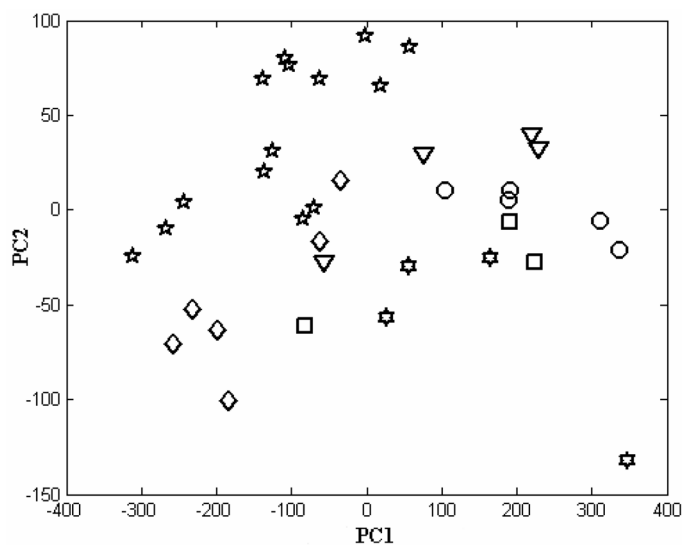


Рис. 1. График счетов молдавских винных дистиллятов возрастом от 4 до 7 лет в двумерном пространстве главных компонент

пониженной размерности с помощью двух видов обучения с учителем – дискриминантного анализа и классификационных деревьев. При этом для дискриминантного анализа будем применять счета двух главных компонент, для построения классификационного дерева – трех.

Линейный дискриминантный анализ (Linear Discriminant Analysis – LDA) [13], или (как еще его часто называют по имени создателя) дискриминантный анализ Фишера, предназначен для поиска линейной комбинации переменных, проекция на которую максимизирует отношение общей дисперсии выборки к сумме дисперсий внутри отдельных классов. В ряде случаев алгоритм LDA может успешно работать и с перекрывающимися классами. Перекрывание границ классов не играет большой роли, пока их центры находятся достаточно далеко друг от друга, чтобы дисперсия выборки и дисперсия классов существенно отличались друг от друга. Естественным обобщением LDA является квадратичный дискриминантный анализ (Quadratic Discriminant analysis – QDA), который использует для разделения классов параболы. Нами были применены LDA и QDA для парной классификации производителей дистиллятов в двумерном пространстве главных компонент (см. рис. 2, на котором представлены примеры правильной дискриминации обоими методами (а) и одним методом (б), а также неправильной дискриминации обоими методами (в)). Установлено, что попарная дискриминация 6 производителей осуществляется методами LDA и QDA для пар 1–3, 1–4, 2–4, 2–6, 3–5, 4–5 и 4–6. Не осуществляется линейная дискриминация для пар 1–2, 1–5, 1–6, 2–3 и 5–6, но для этих случаев действенным является метод QDA. И три пары производителей (2–5, 3–4 и 3–6) не дискриминируются ни LDA, ни QDA. Таким образом, с помощью дискриминантного анализа удается классифицировать 12 пар производителей молдавских винных дистиллятов из 15.

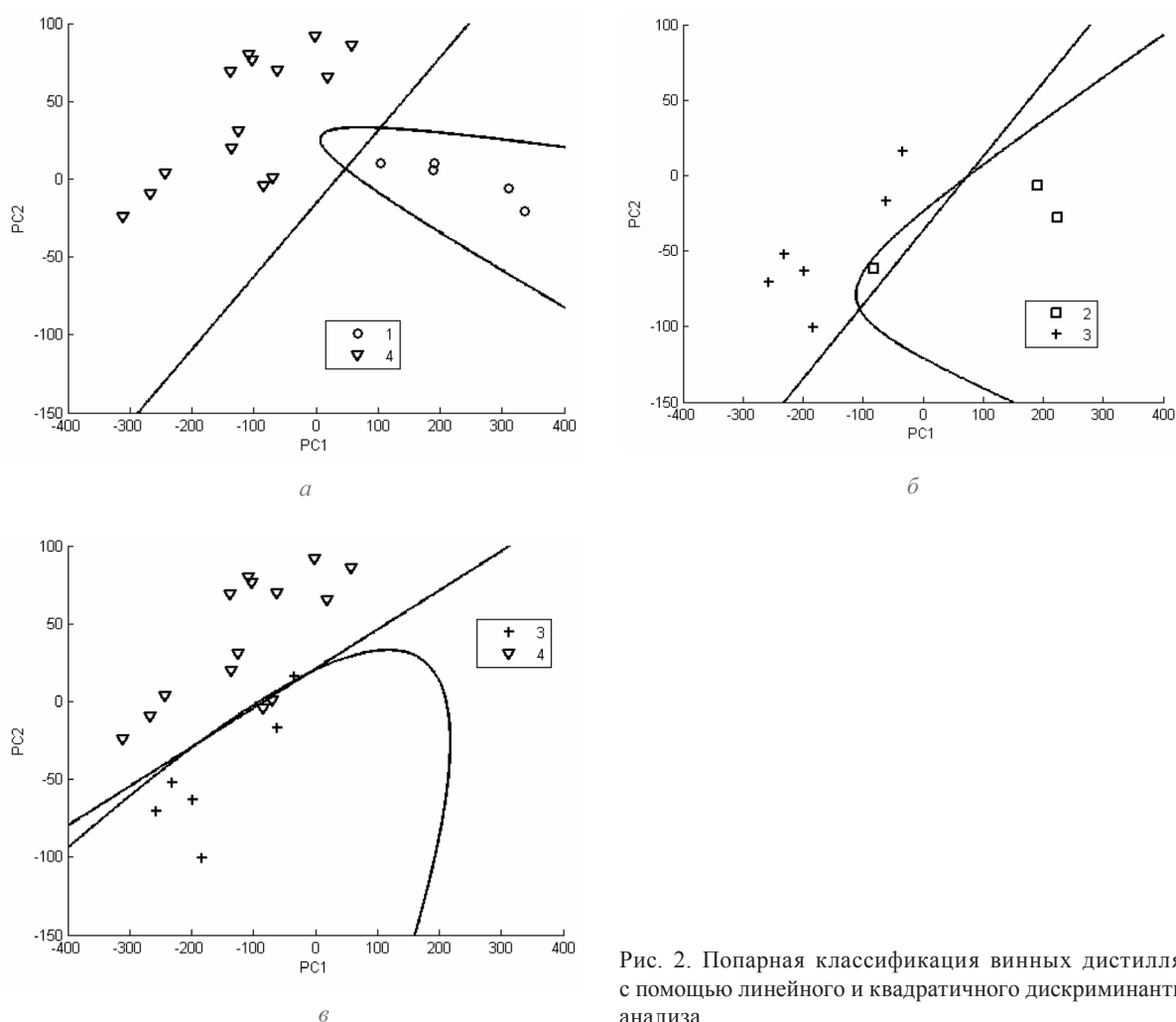


Рис. 2. Попарная классификация винных дистиллятов с помощью линейного и квадратичного дискриминантного анализа

Другим видом обучения с учителем является построение деревьев принятия решений (деревьев классификации или регрессионных деревьев, если прогнозируется количественный показатель) [14]. Деревья классификации представляют собой набор «листьев» и «веток» с бинарными правилами перехода. На «ветках» дерева решения записаны атрибуты, от которых зависит целевая функция, в «листьях» – значения целевой функции. Цель моделирования состоит в том, чтобы предсказать значение целевой переменной на основе нескольких переменных на входе. В нашем случае входными переменными являются три главные компоненты, определенные из спектров пропускания дистиллятов, а целевой переменной – условный номер завода-производителя. Остановка построения дерева классификации происходит в случае, когда каждый «лист» содержит уникальное значение целевой переменной. Наилучший результат построения дерева классификации был получен на основе трех главных компонент при использовании алгоритма, рассматривающего все возможные $2^{L-1}-1$ комбинации построения L -уровневого предиктора, и индекса разнообразия Джини – Симпсона $1 - \sum p_i^2$, где p_i – вероятность того, что дистиллят изготовлен i -м производителем. Величина PC3, меньшая $-0,504488$, определяет принадлежность дистиллята к производителям с номерами 4 или 5. При этом образец относится к производителю 5 при PC2 < $-13,2575$. Величина PC3, большая или равная $-0,504488$, определяет принадлежность дистиллята к производителям с номерами 1, 3 или 6. Среди этих образцов величина PC1 < $-132,251$ классифицирует дистиллят производителя номер 3. Оставшиеся образцы классифицируются величиной PC2. При PC2 < $20,0857$ дистиллят относится к производителю номер 1, в противном случае – к номеру 6. Построенное дерево классификации таким образом способно достоверно определить пять производителей дистиллятов из исследуемых шести.

Вторым признаком аутентичности оригинальности винного дистиллята является его возраст. С помощью метода главных компонент и двух рассмотренных видов классификации на его основе нам не удалось определить возраст дистиллятов исследуемой выборки. Применим для достижения этой цели однооткликтовую регрессию на латентные структуры (PLS1 – projection on latent structures) [15]. Отличие PLS1 от регрессии на главные компоненты состоит в том, что вместо поиска подпространства отсчетов с максимальной корреляцией между предикторами (спектральными отсчетами) и откликами (возраст дистиллятов) PLS1 строит линейную регрессионную модель с помощью одновременного проецирования предикторов и откликов в новое пространство так называемых латентных структур. PLS1 регрессия показывает великолепные результаты в случае, когда матрица отсчетов мультиколлинеарна (спектральные отсчеты на раз-

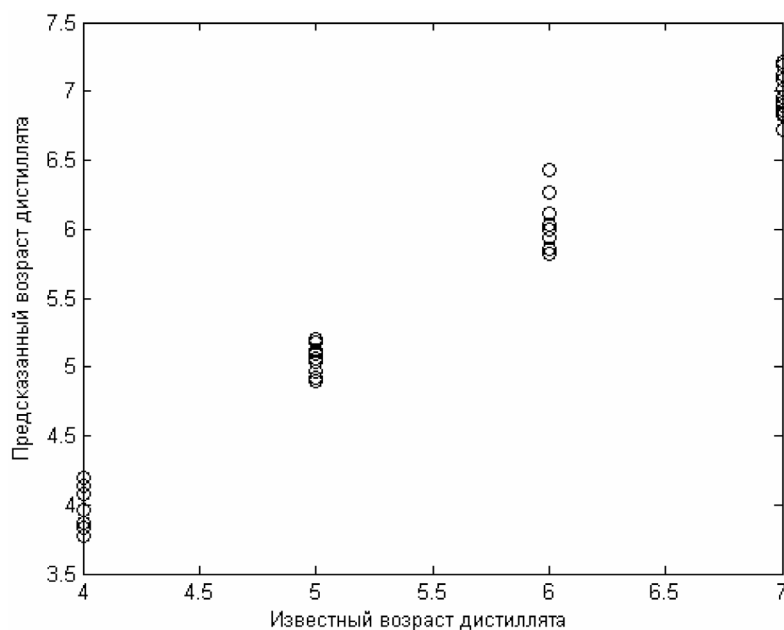


Рис. 3. Результаты применения регрессии на латентные структуры для определения возраста выдержанных винных дистиллятов

ных длинах волн коррелируют друг с другом) и количество образцов значительно уступает количеству спектральных отсчетов. Стандартные методы регрессии в этом случае неэффективны. Именно с таким случаем мы имеем дело. На рис. 3 представлены результаты применения PLS1, полученные для 21 латентной структуры при проведении кросс-валидации, в которой образцы исключались по одному из выборки объемом 42 дистиллята. Параметр регрессии R^2 , представляющий собой разность 1 и отношения суммы квадратов разностей предсказанной и известной величин к сумме квадратов разностей известной и средней величин, равен 0,9801. Видно, что с помощью PLS1 дискретный возраст винных дистиллятов определяется однозначно. Отклонения относительных предсказанных значений возраста от известных показали, что погрешность определения возраста не превышает 8 %.

Итак, применение хемометрических методов (метода главных компонент, дискриминантного анализа, построения деревьев классификации и регрессии на латентные структуры) к широкополосным спектрам пропускания винных дистиллятов позволило определить и производителя, и возраст объектов исследования. Таким образом, продемонстрирован один из возможных способов решения проблемы определения оригинальности выдержанных винных дистиллятов. Метод позволяет установить аутентичность изделий ликеро-водочной промышленности и может быть полезен при разработке средств борьбы с производством контрафактной продукции.

Часть результатов, изложенных в данной работе, была получена в рамках совместных проектов Белорусского республиканского фонда фундаментальных исследований № Ф13МЛД-011 и № 13.820.14.07/ВА Академии наук Молдовы.

Литература

1. Методы контроля, химические и микробиологические факторы. Оценка подлинности и выявление фальсификации молочной продукции: МУ 4.1/4.2.2484-09: утв. Гл. сан. врачом Рос. Федерации 11.02.2009. М., 2009.
2. Куселев В. М., Коркачева О. В., Ганиев Р. М. // Ползуновский вестн. 2012. № 2/2. С. 44–48.
3. Blanco M., Peguero A. // Trends in Analytical Chemistry. 2010. Vol. 29. P. 1127–1136.
4. Gonzaga F. B., Pasquini C. // Analytica Chimica Acta. 2010. Vol. 670. P. 92–97.
5. Xie L. et al. // Food Chemistry. 2009. Vol. 114. P. 1135–1140.
6. Cozzolino D. et al. // Talanta. 2008. Vol. 74. P. 711–716.
7. Kasemsumran S. et al. // Chemometrics and Intelligent Laboratory Systems. 2006. Vol. 82. P. 97–103.
8. Chuah Z. et al. // Chemometrics and Intelligent Laboratory Systems. 2010. Vol. 104. P. 347–351.
9. Massart D. L. et al. Chemometrics: Data Handling in Science and Technology. Elsevier, 1988.
10. Blanco M., Villarroya I. // Trends in analytical chemistry. 2002. Vol. 21. P. 240–250.
11. Скурихин И. М. Химия коньяка и бренди. М., 2005.
12. Abdi H., Williams L. J. // Wiley Interdisciplinary Reviews: Computational Statistics. 2010. Vol. 2. P. 433–459.
13. Fisher R. A. // Annals of Eugenics. 1938. Vol. 8. P. 376–386.
14. Brown S. D., Myles A. J. // Comprehensive Chemometrics: Chemical and Biochemical Data Analysis. 2009. Vol. 3. P. 541–569.
15. Abdi H. // Wiley Interdisciplinary Reviews: Computational Statistics. 2010. Vol. 2. P. 97–106.

N. F. DEGTYAR, E. I. NEZALZOVA, M. V. ROGOVAYA, G. V. SINITSYN, E. A. SKORBANOVA, M. A. KHODASEVICH

IDENTIFICATION OF THE AUTHENTICITY OF MATURED WINE DISTILLATES BY THE UV-, VISIBLE AND NEAR IR-SPECTROSCOPY METHODS

Summary

Ultraviolet, visible and near infra-red spectroscopy is used for identification of the authenticity of matured wine distillates. The use of principal component analysis, the discriminant analysis, the classification trees and projection latent structures to broad-band transmission spectra of wine distillates allowed determining manufacturer and aging of wine distillates.