

УДК 543.632.495; 004.032.26

ИДЕНТИФИКАЦИЯ ОБРАЗЦОВ ВОДЫ ИСТОЧНИКОВ И РЕК г. ХАРЬКОВ: СРАВНЕНИЕ МЕТОДОВ МНОГОМЕРНОГО АНАЛИЗА ДАННЫХ

Я.Н. Пушкарева, А.Б. Следзевская, А.В. Пантелеймонов, Н.П. Титова, О.И. Юрченко,
В.В. Иванов, Ю.В. Холин

(Харьковский национальный университет имени В.Н. Каразина;
e-mail: kholin@univer.kharkov.ua)

Рассмотрено применение искусственных нейронных сетей для установления происхождения образцов вод из разных источников и рек г. Харьков по данным о содержании ионов металлов. На примере образцов речных вод показана возможность верной идентификации образцов даже при наличии пропусков в исходных данных. Изложены рекомендации по выбору оптимального числа нейронов для синтеза нейросетей.

Ключевые слова: качественный химический анализ, идентификация, искусственная нейронная сеть, линейный дискриминантный анализ.

За последние два десятилетия существенно возросла значимость качественного химического анализа, что обусловлено растущей потребностью в определении состава сложных смесей в разных областях: мониторинг состояния окружающей среды, оценка полноты медико-биологических препаратов, идентификация и обеспечение безопасности пищевых продуктов, напитков и пищевого сырья и др. [1]. В связи с этим изменилось и представление о содержании качественного химического анализа. Согласно [2–4], современный качественный анализ позволяет обнаруживать и идентифицировать аналиты, а также проводить дискриминацию образцов. Результатом решения всех этих задач является та или иная классификация объектов анализа: при обнаружении – разделение образцов на группы, содержащие аналит в концентрации, превышающей пороговую, и не содержащие его; при идентификации – заключение о тождественности аналита и эталона или о принадлежности аналита некоторому классу объектов; при дискриминации – разделение массива анализируемых образцов на группы объектов с близкими характеристиками.

Таким образом, все более отчетливой становится тенденция к трактовке качественного анализа как процедуры классификации объектов по их признакам. Зачастую речь идет об обработке массивов многомерных экспериментальных данных, полученных в результате применения хроматографического анализа, различных спектроскопических методов, а также

таких сенсорных систем, как «электронный нос» и «электронный язык». При обработке таких данных неизбежным становится применение современных хемометрических методов – алгоритмов распознавания образов и классификации, дискриминантного анализа, искусственных нейронных сетей и др. Все большее внимание привлекают искусственные нейронные сети, которые успешно применяются в разных отраслях для решения классификационных задач (см., например, [5–9]) благодаря наличию адаптивной архитектуры и способности к обучению.

Одним из инструментов, необходимых для решения комплексной проблемы качества воды, является качественный химический анализ, в особенности идентификация и дискриминация образцов вод по их составу, происхождению, свойствам и т.д. О растущем внимании к проблеме качества воды свидетельствует экспоненциальный рост числа публикаций, реферируемых информационным ресурсом ScienceDirect (<http://www.sciencedirect.com/>) и содержащих в названиях и рефератах словосочетание «quality of water» (рис. 1). При этом хемометрические методы идентификации и оценки качества образцов вод (метод главных компонент, факторный, кластерный и дискриминантный анализ, искусственные нейронные сети и др.) приобретают все большее значение [10–20].

Настоящая статья посвящена применению нейронных сетей для установления происхождения¹ образцов

¹Под установлением происхождения образцов вод понимают их отнесение к тому или иному заранее выбранному классу, т.е. идентификацию образцов вод (ср. с определением: идентификация – отождествление аналита с известным веществом (эталон) или отнесение аналита к группе соединений [4]).

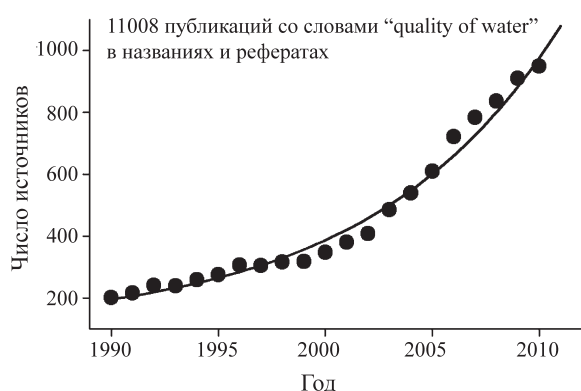


Рис. 1. Данные о публикациях в базе данных ScienceDirect, относящихся к проблеме качества воды (по состоянию на 30.05.2012)

речных (загрязненных промышленными стоками) и родниковых вод г. Харьков на основе данных о содержании ионов 8 металлов. Выбор объектов исследования объясняется изменением химического состава не только поверхностных, но и подземных вод в результате интенсивного антропогенного воздействия.

В работе поставлены следующие цели: изучить эффективность работы нейронных сетей «с обучением» для решения задачи идентификации образцов вод; проверить устойчивость различных алгоритмов нейронных сетей к наличию пропусков в исходных данных о характеристиках образцов вод; сформулировать рекомендации по выбору оптимального числа нейронов для синтеза нейронных сетей.

Для выполнения расчетов использован пакет MATLAB 6.5 [21, 22].

Обрабатываемые данные

Анализируемый массив данных включает 22 образца речных вод и 24 образца родниковых вод из рек и источников г. Харьков, отобранных в разные сезоны в течение 2008–2010 гг. В образцах вод определяли концентрацию ионов меди, цинка, свинца, кадмия, марганца, железа, кобальта и никеля (Приложения 1, 2).

Для каждой реки и источника число отобранных проб (объемом по 500 мл) колебалось от 2 до 7. Время транспортировки и хранения проб от момента их отбора до обработки не превышало суток.

Добавлением концентрированной HNO_3 устанавливали значение pH раствора, равное 2. Затем пробы упаривали до объема 20–22 мл, переносили в мерные колбы вместимостью 25 мл, доводили до метки бидистиллированной водой и тщательно перемешивали. Содержание ионов металлов в образцах вод определяли методом атомно-абсорбционной спектроскопии [23–26] на спектрометре «С-115-М1» («СЕЛМИ»),

Украина) в пламени пропан–бутан–воздух и ацетилен–воздух при оптимальных параметрах определения металлов. Градуировочные растворы, содержащие от 0,1 до 1,0 мг/л ионов металлов, готовили из образцов со стандартным составом раствора (концентрация ионов металлов 1 г/л) путем разбавления исходных водных растворов бидистиллированной водой [27]. Относительное стандартное отклонение определяемых концентраций не превышало 0,01–0,02.

Поскольку концентрация ионов металлов в образцах вод варьировала в диапазоне от тысячных до десятых мг/л, перед применением алгоритмов классификации провели масштабное преобразование исходных данных:

$$x_i^{\text{norm}} = \frac{x_i - \bar{x}}{\text{std}(x)}, i = 1, 2, \dots, N,$$

где x^{norm} – преобразованная безразмерная концентрация данного элемента в i -м образце вод (величина x^{norm} имеет нулевое среднее и единичную дисперсию), x_i – концентрация данного элемента в i -м образце, \bar{x} – среднее значение концентрации данного элемента в образцах, $\text{std}(x)$ – стандартное отклонение концентрации данного элемента в образцах, N – число образцов вод.

Классификация образцов вод с применением алгоритмов «с обучением»

Для установления происхождения образцов вод применили алгоритмы нейронных сетей «с обучением» (вероятностную сеть, динамическую сеть, каскадную сеть, сеть прямого распространения сигнала, сеть Элмана), а также линейный дискриминантный анализ.

Искусственная нейронная сеть – это набор нейронов, соединенных между собой. Синтез нейронной сети происходит в два этапа: выбор архитектуры сети (определение числа слоев, нейронов, типов функций активации) и выбор способа обучения сети (подбор весовых коэффициентов и правила их корректировки). Для многих нейронных сетей этапом, предваряющим процедуру их обучения, является инициализация весов и смещений сети. В данной работе использовали функцию инициализации слоя нейронов по алгоритму Нгуена–Видроу [28]. В качестве метода обучения для каскадной сети, сети с прямым распространением сигнала и динамической сети использовали функцию, реализующую метод обучения, сущность которого заключается в настройке весов и смещений сети с использованием алгоритма Левенберга–Марквардта;

для сети Элмана используют алгоритм обратного распространения ошибки [29, 30]. Критерием остановки обучения сети являлось достижение значения средней квадратичной ошибки обучения, равного 0,01. Подробную информацию об архитектуре исследуемых нейронных сетей и алгоритмов обучения можно найти в работах [21, 31, 32]. Методы управляемого обучения предусматривают наличие обучающей и тестовой выборок. Обучающая выборка (training sample) представляет собой множество объектов с известной принадлежностью к тем или иным классам; эта выборка используется для нахождения оптимальных параметров алгоритма. Тестовая выборка (testing sample) – множество исследуемых объектов с неизвестной принадлежностью к классам, по которому оценивается эффективность работы алгоритма.

Первым шагом применения алгоритмов классификации «с обучением» является определение оптимального состава обучающей выборки. Для формирования обучающей выборки использовали *вероятностную нейронную сеть*, обладающую простой архитектурой, коротким временем обучения, возможностью применения полученных результатов для обучения остальных типов нейронных сетей. Вероятностная нейросеть требует оптимизации единственного параметра – значения степени сглаживания δ (рекомендуют использовать значение δ , равное 0,1 [33]), а число нейронов скрытого слоя равно числу образцов в обучающей выборке.

Результаты качественного химического анализа невозможно оценивать с помощью такой метрологической характеристики, как принятая в количественном анализе «неопределенность» [34]. Вместо «неопределенности» предложено использовать «недостоверность» (unreliability) [35, 36]. Высокой достоверности методики соответствует низкая доля делаемых на основе ее применения ошибочных заключений [37]. В настоящей работе недостаточность идентификации оценивали по доле неправильно классифицированных образцов вод:

$$P = \frac{n}{N} \cdot 100, \%,$$

где n – число неверно классифицированных образцов вод в тестовой выборке, N – общее число образцов в тестовой выборке².

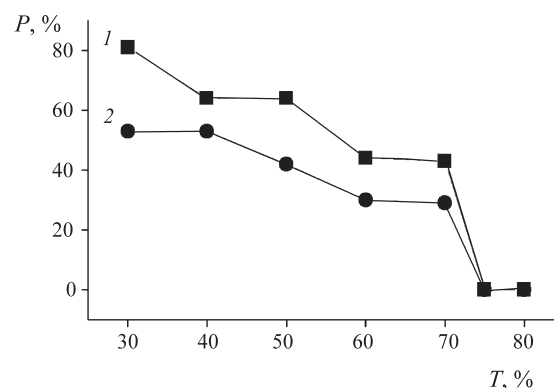


Рис. 2. Зависимость значений долей неправильно классифицированных образцов вод тестовой выборки от объема обучающей выборки для вероятностной нейросети: 1 – образцы речных вод, 2 – образцы родниковых вод

На рис. 2 представлены зависимости долей неправильно классифицированных образцов вод тестовой выборки от объема обучающей выборки (T , %) для речных и родниковых вод (T – доля образцов в обучающей выборке от их общего числа).

Таким образом, для обучения разных типов нейронных сетей использовали объем обучающей выборки 75%, что эквивалентно 16 и 18 образцам речных и родниковых вод соответственно. Параметры обучения и тестирования реализованных нейронных сетей, а также минимальные ошибки классификации образцов вод представлены в табл. 1. Для реализованных нейронных сетей определены оптимальное число нейронов скрытого слоя (число нейронов выходного слоя принималось тождественным числу классов, т.е. числу исследуемых рек или источников) и оптимальная комбинация функций активации для скрытого и выходного слоев. Использование именно таких параметров обеспечивало корректное обучение сети и удовлетворительную идентификацию образцов вод тестовой выборки (кроме сети Элмана).

Оптимальное число нейронов – это число нейронов, при котором наблюдается минимальная недостаточность классификации / идентификации исследуемых образцов. На рис. 3 представлен пример определения оптимального числа нейронов, а именно зависимость долей неправильно классифицированных образцов вод от числа нейронов скрытого слоя для динамической нейронной сети.

При использовании для идентификации образцов вод линейного дискриминантного анализа [38] недостаточность P неприемлемо высока (67 и 33% для

²Подразумевается, что корректно обученная нейронная сеть классифицирует образцы обучающей выборки безошибочно.

Т а б л и ц а 1

Параметры нейронных сетей и значения минимальной недостоверности идентификации образцов вод

Нейронная сеть	Число нейронов для классификации образцов речных / родниковых вод	Функции активации для скрытого / выходного слоев	P, % для образцов речных / родниковых вод
Каскадная сеть	9/10	гиперболический тангенс / гиперболический тангенс	17/0
Сеть с прямым распространением сигнала	9/11	гиперболический тангенс / линейная	17/0
Сеть Элмана	11/12	гиперболический тангенс / линейная	17/17
Динамическая сеть	11/11	гиперболический тангенс / линейная	0/0
Вероятностная сеть	16/18	Гаусса / линейная	0/0

образцов речных и родниковых вод соответственно). Таким образом, для идентификации образцов вод по данным о содержании микроэлементов рекомендуется использовать вероятностную и динамическую нейронные сети.

Устойчивость алгоритмов классификации к наличию пропусков в исходных данных

Для оценки устойчивости алгоритмов классификации к наличию пропусков в исходных данных использовали 4 образца речных вод, анализ которых проводили в 2011 г. и в которых концентрации ионов никеля не были определены (*Приложение 3*).

Для обучения нейросетей использовали 22 образца вод (*Приложение 2*) и предложенные в табл. 1 функции активации. Отсутствующие значения концентрации ионов никеля заполнили средним значением их концентрации в образцах обучающей выборки (0,0144 мг/л). В табл. 2 представлены характеристики идентификации образцов речных вод по данным с пропусками – значения минимальной недостоверности и оптимальное число скрытых нейронов. Наихудшие результаты получены при применении линейного дискриминантного анализа; достоверную идентификацию обеспечивает применение динамической сети. Таким образом, на примере образцов речных вод показана возможность достоверной идентификации образцов вод с применением нейронных сетей даже в случае отсутствия одной из характеристик, описывающих образцы обучающей выборки (эталонов).

Рекомендации по выбору числа нейронов

Анализируя оптимальную архитектуру нейронных сетей (табл. 1, 2), можно прийти к заключению, что

при классификации образцов по данным многооткликового эксперимента оптимальное число нейронов скрытого слоя находится в интервале, определяемом по формуле

$$n_{\text{neurons}} = \left[\frac{n_{\text{training}}}{n_{\text{testing}}} + n_{\text{properties}} \right] \pm \frac{n_{\text{classes}}}{2},$$

где n_{training} – число образцов в обучающей выборке, n_{testing} – число образцов в тестовой выборке, $n_{\text{properties}}$ – число характеристик, n_{classes} – число классов.

Отметим, что предложенный способ определения оптимального числа нейронов является эвристическим и требует проверки при анализе новых экспериментальных данных.

Таким образом, показана эффективность применения аппарата нейронных сетей для установления про-

Т а б л и ц а 2

Результаты идентификации образцов речных вод при наличии в данных пропусков

Алгоритм	Число нейронов	P, %
Каскадная сеть	14	25
Сеть с прямым распространением сигнала	15	25
Сеть Элмана	15	25
Динамическая сеть	14	0
Вероятностная сеть	22	25
Линейный дискриминантный анализ	–	75

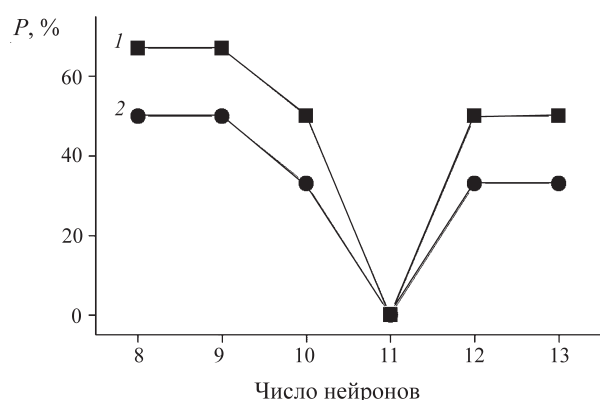


Рис. 3. Зависимость долей неправильно классифицированных образцов вод от числа нейронов для динамической сети: 1 – образцы речных вод, 2 – образцы родниковых вод

исхождения образцов вод по данным о содержании ионов 8 металлов даже в случае отсутствия сведений о концентрации одного из них. Наиболее эффективны для решения такого рода задач динамическая и вероятностная нейронные сети. Приведенные результаты

позволили сформулировать некоторые правила для упрощения процедуры подбора параметров нейронных сетей для решения классификационных задач по многомерным массивам экспериментальных данных:

для реализации нейросетей (кроме вероятностной нейросети) достаточно оперировать двумя типами функций активации – линейной и гиперболическим тангенсом;

оптимальное число нейронов находится в интервале, для расчета которого необходимы только начальные сведения о задаче (число образцов в обучающей и тестовой выборках, число характеристик и число классов);

оптимальный объем обучающей выборки, определенный с помощью вероятностной нейронной сети, можно применять для обучения разных типов нейронных сетей.

Полученные результаты могут быть полезными для выявления ассортиментной фальсификации минеральных вод или для установления соответствия заявленного происхождения воды истинному.

Приложение 1

Концентрация ионов металлов в образцах родниковых вод (мг/л, 10⁻¹)

Металл / Источник	Zn	Cu	Mn	Fe	Cd	Pb	Co	Ni
Саржин Яр «Харьковская-1» (2010 г.)	0,090	0,100	0,300	0,480	0,024	0,250	0,084	0,080
	0,130	0,080	0,240	0,720	0,024	0,280	0,063	0,090
	0,120	0,080	0,260	0,500	0,018	0,250	0,078	0,090
	0,120	0,050	0,200	1,020	0,019	0,230	0,084	0,110
«Харьковская -2» (2010 г.)	0,070	0,050	0,180	0,570	0,012	0,230	0,069	0,120
	0,050	0,040	0,130	0,590	0,016	0,190	0,069	0,110
	0,070	0,040	0,080	0,330	0,018	0,160	0,103	0,120
	0,080	0,050	0,190	0,570	0,021	0,230	0,063	0,120
	0,190	0,180	0,080	0,640	0,029	0,350	0,078	0,200
	0,080	0,750	0,610	0,760	0,024	0,290	0,056	0,200
	0,080	0,230	0,700	1,200	0,022	0,280	0,063	0,200
Пантелеймоновская церковь (2010 г.)	0,080	0,190	0,220	0,550	0,019	0,350	0,094	0,200
	0,100	0,930	0,180	0,370	0,025	0,250	0,078	0,200
	0,050	0,070	0,410	0,280	0,019	0,000	0,078	0,090
	0,060	0,080	0,210	0,280	0,026	0,000	0,063	0,100

Продолжение приложения 1

Металл Источник	Zn	Cu	Mn	Fe	Cd	Pb	Co	Ni
Завод пищевых кислот (2010 г.)	0,100	0,400	0,140	0,260	0,012	0,160	0,078	0,100
	0,140	0,150	0,240	0,780	0,026	0,160	0,078	0,140
	0,230	0,220	0,130	0,430	0,024	0,130	0,100	0,140
ул. Уборевича (2009 г.)	0,230	0,050	0,110	0,080	0,083	0,110	0,133	0,120
	0,140	0,040	0,020	0,150	0,053	0,150	0,125	0,140
	0,180	0,050	0,030	0,150	0,049	0,270	0,063	0,140
	0,140	0,020	0,040	0,100	0,052	0,170	0,000	0,110
Парк «Юность», (2009 г.)	0,340	0,040	0,150	0,520	0,123	0,210	0,125	0,380
	0,150	0,030	0,030	0,290	0,086	0,070	0,125	0,060

Приложение 2

Концентрация ионов металлов в образцах речных вод (мг/л)

Металл Река	Zn	Cu	Mn	Fe	Cd	Pb	Co	Ni
Немышля (2008 г.)	0,0150	0,0055	0,0230	0,1000	0,0021	0,0350	0,0084	0,0200
	0,0130	0,0045	0,0410	0,1200	0,0015	0,0440	0,0088	0,0180
	0,0080	0,0073	0,0170	0,0280	0,0024	0,0350	0,0094	0,0140
	0,0120	0,0073	0,0130	0,0380	0,0022	0,0470	0,0063	0,0130
	0,0120	0,0078	0,4330	0,0190	0,0022	0,0460	0,0140	0,0200
Харьков (2009 г.)	0,0050	0,0034	0,0020	0,0220	0,0017	0,0410	0,0105	0,0160
	0,1040	0,0103	0,0070	0,0190	0,0019	0,0410	0,0112	0,0120
	0,0150	0,0043	0,0030	0,0170	0,0017	0,0340	0,0090	0,0080
	0,0070	0,0052	0,0030	0,0160	0,0039	0,0310	0,0085	0,0100
	0,0100	0,0052	0,0030	0,0170	0,0030	0,0230	0,0097	0,0070
Лопань (2009 г.)	0,0470	0,0043	0,0050	0,0110	0,0032	0,0310	0,0070	0,0120
	0,1090	0,0069	0,0030	0,0170	0,0052	0,0430	0,0068	0,0140
	0,0650	0,0069	0,1330	0,0220	0,0060	0,0380	0,0078	0,0130
	0,0100	0,0078	0,0060	0,0170	0,0030	0,0280	0,0075	0,0190
	0,0110	0,0087	0,0400	0,0230	0,0023	0,0380	0,0107	0,0160
Уды (2010 г.)	0,0120	0,0056	0,2790	0,0690	0,0000	0,0000	0,0125	0,0110
	0,0170	0,0056	0,0030	0,1460	0,0000	0,0000	0,0000	0,0040
	0,0050	0,0032	0,0000	0,0110	0,0000	0,0000	0,0071	0,0040
	0,0300	0,0286	0,0150	0,0970	0,0090	0,2000	0,0554	0,0300
	0,0090	0,0063	0,0230	0,0460	0,0000	0,0380	0,0089	0,0390
	0,0080	0,0110	0,0130	0,0430	0,0021	0,0100	0,0047	0,0070
	0,0110	0,0080	0,0280	0,0460	0,0015	0,0250	0,0078	0,0090

Приложение 3

Концентрация ионов металлов в образцах речных вод, для которых не определена концентрация ионов никеля (мг/л) (2011 г.)

Металл \ Река	Zn	Cu	Mn	Fe	Cd	Pb	Co
Немышля	0,0520	0,0330	0,0960	0,2130	0,0035	0,0250	0,0050
Харьков	0,0340	0,0300	0,0660	0,1800	0,0035	0,0160	0,0070
Лопань	0,0360	0,0340	0,0370	0,1600	0,0035	0,0240	0,0080
Уды	0,0420	0,0330	0,0470	0,0690	0,0031	0,0090	0,0030

СПИСОК ЛИТЕРАТУРЫ

1. Вершинин В.И., Дерендяев Б.Г., Лебедев К.С. Компьютерная идентификация органических соединений. М., 2002.
2. Vlasov Yu., Legin A., Rudnitskaya A., Di Natale C., D'Amico A. // Pure Appl. Chem. 2005. **77**. N 11. P. 1965.
3. Hardcastle W.A. (Editor). Qualitative analysis: a guide to best practice. Cambridge, 1998.
4. Мильман Б.Л. Введение в химическую идентификацию. СПб., 2008.
5. Groselj N., van der Veer G., Tusar M., Vracco M., Novic M. // Food Chem. 2010. **118**. N 4. P. 941.
6. Cevoli C., Cerretani L., Gori A., Caboni M.F., Gallina Toschi T., Fabbri A. // Food Chem. 2011. **129**, N 3. P. 1315.
7. Peres A.M., Baptista P., Malheiro R., Dias L.G., Bento A., Pereira J.A. // Chem. Intell. Lab. Syst. 2011. **105**. N 1. P. 65.
8. Liao S.-H., Wen C.-H. // Expert Syst. Appl. 2007. **32**. N 1. P. 1.
9. Balabin R.M., Safieva R.Z., Lomakina E.I. // Anal. Chim. Acta. 2010. **671**. P. 27.
10. Ouyang Y. // Water Res. 2005. **39**. N 12. P. 2621.
11. Parinet B., Lhote A., Legube B. // Ecol. Model. 2004. **178**. N 3-4. P. 2953.
12. Stanimirova I., Polowniak M., Skorek R., Kita A., John E., Buhl F., Walczak B. // Talanta. 2007. **74**. N 1. P. 153.
13. Panda U.Ch., Sundaray S.K., Rath P., Nayak B.B., Bhatta D. // J. Hydrol. 2006. **331**. N 3-4. P. 434.
14. Rodrigues P.M.S.M., Rodrigues R.M.M., Costa B.H.F., Martins A.A.L.T., da Silva J.C.G.E. // Chem. Intell. Lab. Syst. 2010. **102**. N 2. P. 130.
15. Saim N., Osman R., Spian D.R.S.A., Jaafar M.Z., Juahir H., Abdullah M.P., Ghani F.A. // Water Res. 2009. **4**. N 20. P. 5023.
16. Zhou F., Guo H., Liu Y., Jiang Y. // Mar. Poll. Bull. 2007. **54**. N 6. P. 745.
17. Terrado M., Lavigne M.-P., Tremblay S., Duchesne S., Ville-neuve J.-P., Rousseau A.N., Barcelo D., Tauler R. // J. Hydrol. 2009. **369**. N 3-4. P. 416.
18. Felipe-Sotelo M., Andrade J.M., Carlosena A., Tauler R. // Anal. Chim. Acta. 2007. **583**. N 1. P. 128.
19. Groselj N., van der Veer G., Tusar M., Vracco M., Novic M. // Food Chem. 2010. **118**. N 4. P. 941.
20. Brodnjak-Voncina D., Dobcnik D., Novic M., Zupan J. // Anal. Chim. Acta. 2002. **462**. N 1. P. 87.
21. Дьяконов В.П., Круглов В.В. MATLAB 6.5 SP1/7/7 SP1/7 SP2+Simulink 5/6. Инструменты искусственного интеллекта и биоинформатики. Серия "Библиотека профессионала". М., 2006.
22. Дьяконов В., Круглов В. Математические пакеты расширения MATLAB: Специальный справочник. СПб., М., Харьков, Минск, 2001.
23. Юферова Е.В., Смагунова А.Н., Сизых Ю.И. // Журн. аналит. химии. 1997. **52**. № 9. С. 905.
24. Басаргин Н.Н., Сванидзе З.С., Розовский Ю.Г. // Заводская лаб. 1993. **59**. № 2. С. 8.
25. Юрченко О.И., Тутова Н.П., Козлова О.В. // Вісн. Харк. нац. ун-ту. 2003. № 596. Хімія. Вип. 10 (33). С. 110 (<http://chembull.univer.kharkov.ua/archiv/2003/20.pdf>).
26. Yebra-Biurrun M.C., Bermejo-Barrera A., Bermejo-Barrera M.P., Barciela-Alonso M.C. // Anal. Chim. Acta. 1995. **303**. N 2-3. P. 341.
27. Рубель А.П., Ковальчук Л.И. // Журн. аналит. химии. 1993. **48**. № 5. С. 911.
28. Moody J., Utans J. // Adv. Neur. In. 1992. **4**. P. 683.
29. Fan J., Pan J. // Appl. Math. Comput. 2009. **207**. N 2. P. 351.
30. KrishnaKumar K. // Neurocomputing. 1993. **5**. N 6. P. 273
31. Осовский С. Нейронные сети для обработки информации. М., 2002.
32. Рутковская Д., Пилинський М., Рутковский Л. Нейронные сети, генетические алгоритмы и нечеткие системы. М., 2006.
33. Краснянчин Я.Н., Пантелеймонов А.В., Холин Ю.В. // Вісн. Харк. нац. ун-ту. 2010. № 932. Хімія. Вип. 19 (42). С. 170 (http://chembull.univer.kharkov.ua/archiv/2010_2/21.pdf).
34. Руководство ЕВРАХИМ/СИТАК «Количественное описание неопределенности в аналитических измерениях». СПб., 2002.
35. Rios A., Barcelo D., Buydens L., Cardenas S., Heydorn K., Karlberg B., Klemm K., Lendl B., Milman B., Neidhart B., Stephany R.W., Townshend A., Zschunke A., Valcarcel M. // Accred. Qual. Assur. 2003. **8**. P. 68.
36. Холин Ю.В., Никитина Н.А., Пантелеймонов А.В., Решетняк Е.А., Бугаевский А.А., Логинова Л.П. Метрологические характеристики методик обнаружения с бинарным откликом. Харьков, 2008.
37. Valcarcel M., Cárdenas S., Barcelo D. Metrology of qualitative chemical analysis. Luxembourg, 2002.
38. Park Ch.H., Park H. // Pattern Recogn. 2008. **41**. N 3. P. 1083.

IDENTIFICATION OF WATER SAMPLES FROM DIFFERENT SPRINGS AND RIVERS OF KHARKIV: COMPARISON OF METHODS OF MULTIVARIATE DATA ANALYSIS

Ya.N. Pushkarova, A.B. Sledzevskaya, A.V. Panteleimonov, N.P. Titova, O.I. Yurchenko, V.V. Ivanov, Yu.V. Kholin

Application of artificial neural networks for identification of water samples from different Kharkiv springs and rivers (Eastern Ukraine) was studied. The initial experimental data consisted of metal ions concentrations in the water samples. The artificial neural networks algorithms were demonstrated to provide the correct identification of water samples even in the presence of gaps in the initial data. To realize successfully each particular neural network it is necessary to determine the optimal number of neurons. The corresponding rule was proposed.

Key words: *qualitative chemical analysis, identification, artificial neural network, linear discriminant analysis.*

Сведения об авторах (сотрудники химического факультета Харьковского национального университета имени В.Н. Каразина, Украина, 61022, г. Харьков, пл. Свободы, 4): *Пушкарёва Ярослава Николаевна* – аспирант, ассистент кафедры химического материаловедения (yagoslava.kr@mail.ru); *Следзевская Анастасия Богдановна* – студент, инженер кафедры химического материаловедения (sledzevska@iua); *Пантелеймонов Антон Витальевич* – доцент кафедры химического материаловедения, канд. хим. наук (panteleimonov@univer.kharkov.ua); *Титова Надежда Петровна* – ст. науч. сотр. кафедры химической метрологии; *Юрченко Олег Иванович* – профессор кафедры химической метрологии, докт. хим. наук (yurchenko@univer.kharkov.ua); *Иванов Владимир Венедиктович* – профессор кафедры химического материаловедения, докт. хим. наук (vivanov@univer.kharkov.ua); *Холин Юрий Валентинович* – профессор кафедры химического материаловедения, докт. хим. наук (kholin@univer.kharkov.ua).