

6. Патент на изобретение РФ №2414141. Способ производства кондитерского изделия на основе топинамбура типа нуги и его состав [Текст] / М.И. Дождалева (Курлаева); опубл. 20.03.2011 г., Бюл. № 8.

УДК 664.66: 664.66.016

**ШМАЛЬКО Н.А., канд. техн. наук, доцент, РОСЛЯКОВ Ю.Ф., д-р техн. наук, профессор,  
РУДЕНКО О.В., соискатель, ТОКАРЕВА Ю.А., студент**  
ГОУ ВПО "Кубанский государственный технологический университет", Краснодар, Россия  
**ТЕХНОЛОГИЯ DATA MINING ДЛЯ КОМПЛЕКСНОЙ ОЦЕНКИ  
КАЧЕСТВА ХЛЕБОПЕКАРНОГО СЫРЬЯ, ПОЛУФАБРИКАТОВ  
И ГОТОВОЙ ПРОДУКЦИИ**

В статье приводятся результаты практической реализации технологии Data Mining, в основу которой положен нейросетевой метод проведения статистических процедур для оценки качества хлебопекарного сырья, полуфабрикатов и готовой продукции в случае введения обогатителя – амарантовой муки. Процедура статистической обработки выборки совокупности технологических параметров вначале включала проведение классического факторного анализа с последующим построением обобщённо-регрессионной нейронной сети (GRNN), позволяющей предсказывать значение показателя состояния изучаемой параметрической схемы производства обогатённого пшеничного хлеба – его балльную оценку качества.

**Ключевые слова:** Хлебопекарное производство, комплексная оценка качества, технология Data Mining.

The article presents the results of the practical implementation of the technology Data Mining, one based on neural network method of statistical procedures for assessing the quality of bakery raw materials, semi finished and finished products in case of introduction of fortifier - amaranth flour. The procedure of statistical processing of the sample collection process parameters initially included a classical factor analysis, followed by the construction of a generalized regression neural network (GRNN), which allows to predict the value indicator of the studied parametric scheme of production of enriched white bread - his scoring qualities.

**Keywords:** Bakery production, integrated assessment of quality, technology, Data Mining.

Как известно, качество хлебопекарной продукции определяется целой совокупностью технологических процессов, протекающих во времени сопряженно и взаимосвязано. Так, метаболизм бродильной микрофлоры зависит от состава тестовых полуфабрикатов, в котором он протекает и обусловлен ферментативным гидролизом. Наряду с этим, процессы набухания, растворения и деформации тестовых полуфабрикатов в значительной степени определяются ферментативным гидролизом и метаболизмом бродильной микрофлоры. И наоборот, ферментативный гидролиз, оказывая влияние на другие процессы, сам обусловлен наличием в среде продуктов метаболизма микрофлоры, степенью ее набухания и величиной деформации при замешивании полуфабрикатов [1]. Для формализации технологического процесса хлебопекарного производства эффективным является проведение его многоуровневой структуризации от элементарных актов до технологической системы, где наиболее управляемым в случае внедрения в рецептуру обогатителя является процессор, представленный элементарным объемом субстрата, в котором протекают технологические операции, обусловленные процессами набухания, деформации, ферментативного гидролиза и метаболизма бродильной микрофлоры [2]. Так, в литературных источниках приводится пример программной реализации параметрических схем хлебопекарного производства, которая сводится к выведению суммарного критерия оптимальности [3], что явно не достаточно для формализации совокупности технологических процессов. Применение квалиметрических методов («весовой» метод, метод функции желательности и др.), применяемых для оптимизации процесса по нескольким критериям, также является не эффективным в силу заложенного в перечисленных методах конфликта между критериями: повышение эффективности процесса по одному критерию оптимальности вызывает ухудшение результата про-

цесса по другому критерию [4].

В последнее время в пищевой технологии для решения задач моделирования технологических процессов производства "сложных" систем применяют искусственные нейронные сети - вычислительные структуры, моделирующие биологические свойства, ассоциируемые с процессами человеческого мозга. Адаптируемые и обучаемые, они представляют собой распараллеленные системы, способные к обучению путем анализа положительных и отрицательных воздействий. С помощью искусственных нейронных сетей решается целый спектр задач, в том числе прогнозирования, классификации, оптимизации и управления технологическими параметрами [5].

Среди достоинств нейросетевого механизма является возможность прогнозирования будущих значений переменных по уже имеющимся значениям этих же или других переменных, предварительно осуществляя процесс так называемого обучения на основе имеющихся данных, что получило распространение в универсальной технологии накопления, обработки и анализа статистических данных Data Mining (с англ. "извлечение информации" или "добыча данных"). Технология Data Mining представляет собой процесс анализа данных, начиная с формирования выборки, проведения разведочного анализа, обработки и трансформации данных, задания и применения модели или вида анализа. Задача такой технологии состоит в выявлении скрытых правил и закономерностей в больших объемах выборки путем проведения автоматического поиска шаблонов (паттернов), характерных для каких-либо фрагментов неоднородных многомерных данных [6].

В качестве объекта исследования являлась выборка экспериментальных данных (табл. 1), полученная при проведении инструментальных методов качества сырья, полуфабрикатов и готовых хлебобулочных изделий в случае внесения в рецептуру пшеничного хлеба различных дозировок амарантовой муки [7]. Реализация статистических процедур осуществлялась с помощью пакета прикладных программ STATISTICA for Windows 6.0 [8, 9].

Для классификации переменных применяли факторный анализ, главной целью которого является сокращение числа переменных (редукция данных) и определение структуры взаимосвязей между ними, тем самым оптимизируя их число в корреляционной матрице. Сокращение достигается путем выделения скрытых общих факторов, объясняющих связи между наблюдаемыми признаками (переменными) объекта, то есть вместо исходного набора переменных появляется возможность анализировать данные по выделенным факторам, число которых значительно меньше исходного числа взаимосвязанных переменных [8].

В начале процедуры построили корреляционную матрицу (табл. 2), анализ которой позволяет оценить степень коррелированности переменных между собой.

Таблиця 1

Выборка экспериментальных данных по качеству сырья, полуфабрикатов и готовых хлебобулочных изделий [7]

DD, %	SK, %	IDK, U	PTS, U	UTS, U	GOS, ml	OFX, ml	UOX, ml/100g	FPI, HD	TKM, G	WM, %	PM, %	OSM, UP	PLM, UP	UPM, UP	BOX, U
0,0	31,6	76,0	7,21	6,62	892,0	900,0	247,0	0,44	2,60	42,0	71,0	44,0	26,0	18,0	78,0
3,0	31,5	75,8	6,16	5,60	934,0	940,0	261,0	0,46	2,60	42,6	73,0	62,0	43,0	19,0	86,6
5,0	31,2	74,5	5,13	4,55	1024,0	965,0	268,0	0,48	2,60	42,8	73,0	69,0	49,0	20,0	89,0
7,0	30,9	72,6	4,96	4,39	1072,0	990,0	275,0	0,49	2,70	43,0	74,0	74,0	52,0	22,0	90,5
10,0	30,8	70,3	4,77	4,22	1116,0	995,0	276,0	0,50	2,70	43,2	74,0	79,0	54,0	25,0	89,0
15,0	30,4	68,5	4,72	4,18	1124,0	1000,0	278,0	0,51	2,70	43,4	75,0	82,0	53,0	29,0	87,0

Условные обозначения:

**INPUT** (независимые переменные): DD, % – дозировка амарантовой муки; SK, % – дозировка сырой клейковины; IDK, U – сопротивление клейковины деформирующей нагрузке сжатия; PTS, U – пластичность теста; UTS, U – упругость теста; GOS, ml – газообразующая способность муки; OFX, ml – объем формового хлеба; UOX, ml/100 g – удельный объем хлеба; FPI, HD – формоустойчивость подовых изделий; TKM, GR – титруемая кислотность мякиша; WM, % – влажность мякиша; PM, % – пористость мякиша; OSM, UP – общая сжимаемость мякиша; PLM, UP – пластичность мякиша; **OUTPUT** (зависимая переменная): BOX, U – балльная оценка хлеба.

В полученной матрице выделяли факторы, для чего в режиме *Factor Analysis Results* задавали следующие показатели: число анализируемых переменных - 16; метод анализа – метод главных компонент (*Principal components*); десятичный логарифм детерминанта корреляционной матрицы; число выделенных факторов – 2; собственные значения 14,3440, 1,23803. В результате были получены предварительные факторные нагрузки, как без вращения, так и к которым применили повороты осей, достигаемые вращением факторов методом варимакс исходных (*Varimax row*).

Метод варимакс позволяет максимизировать дисперсии квадратов исходных факторных нагрузок по переменным для каждого фактора, что эквивалентно максимизации дисперсий в столбцах матрицы квадратов исходных факторных нагрузок. Результаты анализа факторных нагрузок приведены в табл. 3. Приведенные статистические данные указывают на сложность выделения главных факторов, что под-

емыми проведенным факторным анализом. В связи с этим, целесообразным посчитали применить к исследуемой выборке нейросетевой метод, позволяющий установить взаимосвязь между параметрами в случае наличия сложных нелинейных зависимостей [9]. В ходе проведения статистических процедур была построена обобщенно-регрессионная нейронная сеть, параметры которой представлены в табл. 4.

Полученная нейронная сеть относится к обобщенно-регрессионным нейронным сетям (GRNN), предназначенным для решения задач регрессии. При построении таких нейронных сетей учитывается предположение, что наблюдение расположено в данной точке пространства и в этой точке имеется некоторая плотность вероятности. Кластеры из близко лежащих точек указывают на то, что в этом месте плотность вероятности большая. Вблизи наблюдения имеется большее доверие к уровню плотности, а по мере отдаления от него доверие убывает и стремится к нулю.

Таблиця 2

Корреляционная матрица

Переменная	Корреляции (Factor), построчное удаление ПД, N=6														
	DD, %	SK, %	IDK, U	PTS, U	UTS, U	GOS, ml	OFX, ml	UOX, ml/100g	FPI, HD	TKM, GR	WM, %	PM, %	OSM, UP	PLM, UP	UPM, UP
DD, %	1,00	-0,98	-0,98	-0,85	-0,84	0,93	0,89	0,88	0,95	0,82	0,94	0,93	0,91	0,80	0,99
SK, %	-0,98	1,00	0,98	0,87	0,86	-0,95	-0,91	-0,90	-0,96	-0,88	-0,93	-0,91	-0,90	-0,80	-0,97
IDK, U	-0,98	0,98	1,00	0,80	0,79	-0,93	-0,86	-0,84	-0,93	-0,89	-0,90	-0,86	-0,86	-0,73	-0,99
PTS, U	-0,85	0,87	0,80	1,00	1,00	-0,96	-0,98	-0,99	-0,97	-0,74	-0,96	-0,92	-0,98	-0,98	-0,76
UTS, U	-0,84	0,86	0,79	1,00	1,00	-0,95	-0,98	-0,98	-0,96	-0,74	-0,96	-0,91	-0,98	-0,98	-0,75
GOS, ml	0,93	-0,95	-0,93	-0,96	-0,95	1,00	0,98	0,96	0,99	0,88	0,96	0,91	0,96	0,90	0,88
OFX, ml	0,89	-0,91	-0,86	-0,98	-0,98	0,98	1,00	1,00	0,98	0,84	0,98	0,96	0,99	0,97	0,82
UOX, ml/100g	0,88	-0,90	-0,84	-0,99	-0,98	0,96	1,00	1,00	0,98	0,82	0,98	0,96	0,99	0,98	0,81
FPI, HD	0,95	-0,96	-0,93	-0,97	-0,96	0,99	0,98	0,98	1,00	0,84	0,99	0,95	0,98	0,93	0,90
TKM, GR	0,82	-0,88	-0,89	-0,74	-0,74	0,88	0,84	0,82	0,84	1,00	0,81	0,80	0,79	0,70	0,83
WM, %	0,94	-0,93	-0,90	-0,96	-0,96	0,96	0,98	0,98	0,99	0,81	1,00	0,98	1,00	0,95	0,89
PM, %	0,93	-0,91	-0,86	-0,92	-0,91	0,91	0,96	0,96	0,95	0,80	0,98	1,00	0,97	0,93	0,87
OSM, UP	0,91	-0,90	-0,86	-0,98	-0,98	0,96	0,99	0,99	0,98	0,79	1,00	0,97	1,00	0,98	0,84
PLM, UP	0,80	-0,80	-0,73	-0,98	-0,98	0,90	0,97	0,98	0,93	0,70	0,95	0,93	0,98	1,00	0,70
UPM, UP	0,99	-0,97	-0,99	-0,76	-0,75	0,88	0,82	0,81	0,90	0,83	0,89	0,87	0,84	0,70	1,00

твердило проведение графического анализа путем построения графика нагрузок. С учетом выделяемых из матрицы 2 главных факторов, на графике было представлено до 10 групп переменных, что свидетельствует о наличии сложных взаимосвязей между параметрами, до конца не обнаружива-

В методе ядерных оценок в точке, соответствующей каждому наблюдению, помещается некоторая простая функция, затем все они складываются, и в результате получается оценка для общей плотности вероятности. Чаще всего в качестве ядерных функций берутся гауссовы функции

Таблиця 3

Факторные нагрузки

Переменные	Факторные нагрузки (без вращения) Выделение: Главные компоненты (Отмечены нагрузки > ,700000)		Факторные нагрузки (варимакс исходный) Выделение: Главные компоненты (Отмечены нагрузки > ,700000)	
	Фактор 1	Фактор 2	Фактор 1	Фактор 2
DD, %	-0,94236	0,294248	0,878202	0,450988
SK, %	0,95175	-0,283646	-0,877460	-0,465127
IDK, U	0,91244	-0,402463	-0,932751	-0,352854
PTS, U	0,967474	0,221069	-0,534766	-0,835998
UTS, U	0,96448	0,228383	-0,527503	-0,839114
GOS, ml	-0,98347	0,055876	0,740367	0,649765
OFX, ml	-0,99188	-0,107864	0,631546	0,772400
UOX, ml/100 g	-0,98801	-0,148119	0,600565	0,798392
FPI, HD	-0,99687	0,032520	0,733542	0,675811
TKM, GR	-0,85308	0,286177	0,808889	0,394139
WM, %	-0,99337	-0,023690	0,691631	0,713431
PM, %	-0,96940	-0,025546	0,673240	0,697948
OSM, UP	-0,98075	-0,115678	0,624554	0,776483
PLM, UP	-0,94397	-0,323100	0,446467	0,892264
UPM, UP	-0,88826	0,439811	0,941700	0,309271
BOX, U	-0,78513	-0,613119	0,129864	0,987666
Общая дисперсия	14,34397	1,238032	7,900160	7,681845
Доля общего	0,89650	0,077377	0,493760	0,480115

с формой колокола).

Если обучающих примеров достаточно количество, то такой метод дает достаточно хорошее приближение к истинной плотности вероятности. Иными словами, в точку расположения каждого обучающего наблюдения помещается гауссова ядерная функция. Каждое наблюдение свидетельствует о некоторой уверенности в том, что поверхность отклика в данной точке имеет определенную высоту, и эта уверенность убывает при отходе в сторону от точки. GRNN-сеть копирует внутрь себя все обучающие наблюдения и использует их для оценки отклика в произвольной точке. Окончательная выходная оценка сети получается как взвешенное среднее выходов по всем обучающим наблюдениям, где величины весов отражают расстояние от этих наблюдений до той точки, в которой производится оценивание (и, таким образом, более близкие точки вносят больший вклад в оценку). Архитектура данной сети представлена на рис. 1.

Для обучения GRNN-сеть модифицировали таким образом, чтобы радиальные элементы соответствовали не отдельным обучающим случаям, а их кластерам, что уменьшает размеры сети и увеличивает скорость ее обучения. Алгоритм К-средних служит для присвоения начальных значений центрам первого промежуточного слоя сети в предположении, что он состоит из радиальных элементов (Moody and Darkin, 1989; Bishop,

Таблица 4

Параметры обобщенно-регрессионной нейронной сети

Параметры	BOX, U.1	BOX, U.2	BOX, U.3	BOX, U.4	BOX, U.5
Среднее данных	86,68333	86,68333	86,68333	86,68333	<b>86,68333</b>
Стандартное отклонение данных	4,09895	4,09895	4,09895	4,09895	4,09895
Среднее ошибки	0,46896	-0,65002	-0,66043	0,14963	0,04600
Стандартное отклонение ошибки	2,71074	0,95540	1,00152	0,41084	0,04056
Среднее абсолютной ошибки	2,53078	0,65939	0,89270	0,35305	0,05046
Отношение стандартного отклонения	0,66133	0,23308	0,24434	0,10023	0,00989
Корреляция	0,75407	0,97455	0,97416	0,99621	0,99995

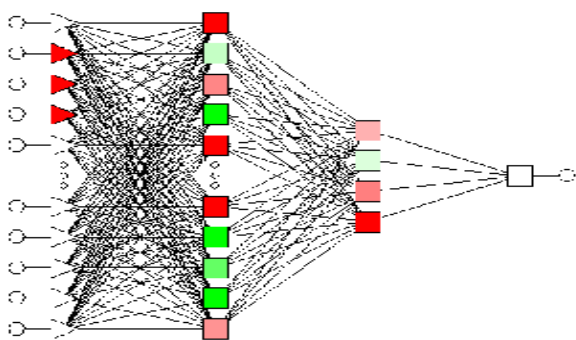


Рис. 1. Архитектура нейронной сети GRNN

1995). В результате данной процедуры проведен регрессионный анализ с коэффициентом корреляции 0,99995 и ошибкой прогнозирования искомого выходного параметра – балльной оценки качества хлеба не более 0,05 %.

Таким образом, применение технологии Data Mining на основе нейросетевого метода позволяет оценить совокупность показателей оценки качества хлебопекарного сырья, полуфабрикатов и готовой продукции с возможностью прогнозирования значения выходного параметра – балльной оценки качества пшеничного хлеба.

Поступила 08.2011

СПИСОК ЛИТЕРАТУРЫ

1. Злобин, Л.А. Оптимизация технологических процессов хлебопекарного производства [Текст] / Л.А. Злобин. – М.: Агропромиздат, 1987. – 200 с.
2. Шмалько, Н.А. Применение структурного анализа для формализации технологического процесса производства функциональных хлебобулочных изделий [Текст] / Н.А. Шмалько // Современные проблемы науки и образования, 2009. – № 6. – С. 14.
3. Краснов, А.Е. Рецептурная модель коллоидных процессов и её программная реализация [Текст] / А.Е. Краснов, Д.Л. Злобин // Хранение и переработка сельхозсырья. – 2005. – № 4. – С. 32-33.
4. Грачев, Ю.П. Математические методы планирования эксперимента [Текст] / Ю.П. Грачев, Ю.М. Плаксин. – М.: ДеЛи принт, 2005. – 296 с.
5. Храмов, А.Г. Использование методики нейронных сетей в пищевой биотехнологии [Текст] / А.Г. Храмов, В.В. Садовой, В.А. Самылина // Известия вузов. Пищевая технология. – 2004. – № 5-6. – С. 105-108.
6. Большаков, П.С. Возможности использования Statistica Data Miner [Текст] / П.С. Большаков // Математика в приложениях. – 2003. – № 1 (1). – С. 13-16.
7. Шмалько, Н.А. Разработка технологий хлебобулочных изделий функционального назначения с использованием продуктов переработки семян амаранта [Текст]: дис. ... канд. техн. наук. – Краснодар, 2005. – 150 с.
8. STATISTICA 6. Статистический анализ данных [Текст]. – 3-е изд. – М.: ООО «Бином-пресс», 2007 – 512 с.: ил.
9. Нейронные сети. STATISTICA Neural Networks: Методология и технологии современного анализа данных [Текст] / Под редакцией В.П. Боровикова. – 2-е изд., перераб. и доп. – М.: Горячая линия – Телеком, 2008. – 392 с.: ил.