

**КОСВЕННАЯ ИДЕНТИФИКАЦИЯ СЕЛЕКЦИОННО-ЗНАЧИМЫХ
ОСОБЕННОСТЕЙ ГЕНОТИПА ПОДСОЛНЕЧНИКА
С ПРИМЕНЕНИЕМ АВТОМАТИЗИРОВАННОГО СИСТЕМНО-
КОГНИТИВНОГО АНАЛИЗА**

Луценко Е. В. – д. э. н., к. т. н., профессор

Кубанский государственный аграрный университет

В статье рассматривается применение АСК-анализа для косвенной (опосредованной) идентификации особенностей генотипа подсолнечника, значимых для селекционной работы, путем исследования его непосредственно наблюдаемых, т.е. наблюдаемых без использования инструментальных средств, фенотипических признаков. Это позволяет сэкономить время и другие ресурсы при отборе растений для следующих селекционных поколений, ускорить процессы селекции и повысить ее качество.

Традиционно отбор растений для формирования следующих селекционных поколений осуществляется на основе инструментального измерения их хозяйственных свойств или исследования генотипа, что может осуществляться лишь в лабораторных условиях, дорого, трудоемко и требует больших затрат времени.

Данная проблема решается путем разработки полевой неинструментальной экспресс-методики, обеспечивающей непосредственно на основе внешне наблюдаемых признаков растений прогнозирование того, будет ли данное растение обладать заданными хозяйственными свойствами.

Идея решения проблемы основана на том, что генотип растения определяет не только его хозяйственные свойства, но фенотип. Поэтому фенотип связан с хозяйственными свойствами. Зная эту связь возможно про-

гнозировать хозяйственные свойства растения по внешне-наблюдаемым фенотипическим признакам.

Необходимо отметить, что на фенотипические признаки, кроме гено-типа растения, оказывают влияние также и внешние для растения факторы. Эти внешние факторы можно разделить на две основные группы по степени их зависимости от воли человека:

1. *Факторы окружающей среды* (прежде всего виды почв и метеорологические факторы) практически не зависят от человека.

2. *Технологические факторы*, т.е. связанные с использованием различных агротехнологий (вспашка, нормы высева, способы удобрения и защиты растений, полив, освещение, севооборот, и т.д.) во многом зависят от человека.

Исследование влияния почв и метеорологических факторов на количественные и качественные результаты выращивания плодовых культур проводились И. А. Драгавцевой, Е. В. Луценко и Л. М. Лопатиной [1–6; 9; 10]. По семечковым, в частности яблокам, подобная работа, насколько известно по литературным данным, по-видимому, впервые проведена А. М. Максимовым [13] под научным руководством Е. В. Луценко. Влияние технологических факторов на количественные и качественные результаты выращивания зерновых колосовых изучалось О. А. Засухиной и Е. В. Луценко [11; 12].

Способ учета влияния всех этих внешних факторов один и тот же и не отличается от способа выявления зависимостей между фенотипом и потребительскими свойствами растений. Это означает, что технически мы могли бы исследовать все эти группы факторов в комплексе. Однако в данном исследовании учитывать мы этого делать не будем по двум основным причинам:

- *первое*: в процессах многолетней селекции подсолнечника морозоустойчивого сорта «Победа» в агрономических журналах *не фиксировались* внешние факторы;

- *второе*: внешние факторы выращивания *не менялись* в процессе селекции, т.е. выращивание осуществлялось в одной микроне с постоянными метеоусловиями и по одной технологии. Поэтому даже если бы эти факторы и учитывались в журналах, их влияние на потребительские свойства изучить не представлялось бы возможным из-за практически полного отсутствия варибельности по этим факторам.

Поэтому в данном исследовании внешними факторами вполне корректно можно пренебречь.

Актуальность для науки данной работы определяется также ее *научной новизной*. Как уже упоминалось выше, в какой-то мере сходные исследования и разработки проводились И. А. Драгавцевой, Л. М. Лопатиной, Е. В. Луценко, А. М. Максимовым, О. А. Засухиной. Однако исследование взаимосвязи фенотипа и потребительских свойств подсолнечника, в частности для морозоустойчивого сорта «Победа», с целями совершенствования методов его компьютерной селекции, по всей видимости проводится *впервые* [8], в этом и состоит *научная новизна данной работы*.

Актуальность для практики темы работы определяется возможностью применения ее результатов *в практике работы* научно-селекционных и образовательных организаций.

В научно-селекционных организациях разработанные методы компьютерной селекции позволяют существенно снизить затраты различных видов ресурсов, прежде всего времени и финансовых средств, на проведение селекционной работы методом отбора лучших по генотипу растений и использования их для формирования следующего селекционного поколения.

В учебном процессе методика разработки и решения перечисленных выше задач может быть использована для разработки полноценной лабораторной работы по дисциплине: "Интеллектуальные информационные системы", изучаемой на 5-м курсе специальности 351400 – Прикладная информатика.

Таким образом, *объектом исследования* является изучение взаимосвязи фенотипа и потребительских свойств культурных растений.

Предмет исследования состоит в изучении взаимосвязи фенотипа и потребительских свойств подсолнечника сорта «Победа».

Цель исследования: выявление причинно-следственных зависимостей между фенотипом и хозяйственными свойствами подсолнечника, и, на основе этого, разработка неинструментальной полевой методики прогнозирования хозяйственных свойств и поддержки принятия решений по отбору растений для следующего селекционного поколения на основе исследования их внешне наблюдаемых признаков.

Данная цель достигается путем декомпозиции в следующую последовательность задач и их решения:

- задача 1: *Выявление* причинно-следственных зависимостей между фенотипическими признаками подсолнечника и его хозяйственными свойствами;

- задача 2: Разработка методики *прогнозирования* хозяйственных свойств растений подсолнечника на основе анализа их фенотипических признаков;

- задача 3: Разработка методики *поддержки принятия решений* по отбору растений для следующего селекционного поколения не по их хозяйственным свойствам, а на основе анализа внешне наблюдаемых фенотипических признаков

Информационной базой для выполнения данной работы являются данные, предоставленные научно-селекционной агрофирмой ООО «Эве-

рест-Олвик», специализирующейся на селекции морозоустойчивых высокотехнологичных сортов подсолнечника, прежде всего совершенствовании сорта «Победа».

Агрофирму ООО «Эверест-Олвик» создал и уже более 13 лет беспрерывно возглавляет известный ученый, продолжатель дела академика В. С. Пустовойта, ученик академика И. А. Рапопорта, заслуженный деятель науки Кубани, профессор, доктор сельскохозяйственных наук Ашот Андраникович Калайджян – автор 225 научных работ, 3-х монографий и двух новых сортов подсолнечника Салют и Победа. Сегодня возглавляемая им агрофирма – одно из ведущих хозяйств России по выведению новых сортов подсолнечника.

Первичная форма существования исходных данных – это полевые журналы агрономов-селекционеров, в которых они на протяжении 35 лет проведения селекционной работы по сорту «Победа» фиксировали регенеративные (потребительские) и вегетативные (фенотипические) характеристики растений на различных делянках. Вид агрономических журналов с исходными данными представлен ниже.

1994			1995			1996			1997			1998		
№	д	г	№	д	г	№	д	г	№	д	г	№	д	г
104	21	94	105	20	95	106	19	96	107	18	97	108	17	98
105	20	94	106	19	95	107	18	96	108	17	97	109	16	98
106	19	94	107	18	95	108	17	96	109	16	97	110	15	98
107	18	94	108	17	95	109	16	96	110	15	97	111	14	98
108	17	94	109	16	95	110	15	96	111	14	97	112	13	98
109	16	94	110	15	95	111	14	96	112	13	97	113	12	98
110	15	94	111	14	95	112	13	96	113	12	97	114	11	98
111	14	94	112	13	95	113	12	96	114	11	97	115	10	98
112	13	94	113	12	95	114	11	96	115	10	97	116	9	98
113	12	94	114	11	95	115	10	96	116	9	97	117	8	98
114	11	94	115	10	95	116	9	96	117	8	97	118	7	98
115	10	94	116	9	95	117	8	96	118	7	97	119	6	98
116	9	94	117	8	95	118	7	96	119	6	97	120	5	98
117	8	94	118	7	95	119	6	96	120	5	97	121	4	98
118	7	94	119	6	95	120	5	96	121	4	97	122	3	98
119	6	94	120	5	95	121	4	96	122	3	97	123	2	98
120	5	94	121	4	95	122	3	96	123	2	97	124	1	98
121	4	94	122	3	95	123	2	96	124	1	97	125	0	98
122	3	94	123	2	95	124	1	96	125	0	97	126	0	98
123	2	94	124	1	95	125	0	96	126	0	97	127	0	98
124	1	94	125	0	95	126	0	96	127	0	97	128	0	98
125	0	94	126	0	95	127	0	96	128	0	97	129	0	98
126	0	94	127	0	95	128	0	96	129	0	97	130	0	98
127	0	94	128	0	95	129	0	96	130	0	97	131	0	98
128	0	94	129	0	95	130	0	96	131	0	97	132	0	98
129	0	94	130	0	95	131	0	96	132	0	97	133	0	98
130	0	94	131	0	95	132	0	96	133	0	97	134	0	98
131	0	94	132	0	95	133	0	96	134	0	97	135	0	98
132	0	94	133	0	95	134	0	96	135	0	97	136	0	98
133	0	94	134	0	95	135	0	96	136	0	97	137	0	98
134	0	94	135	0	95	136	0	96	137	0	97	138	0	98
135	0	94	136	0	95	137	0	96	138	0	97	139	0	98
136	0	94	137	0	95	138	0	96	139	0	97	140	0	98
137	0	94	138	0	95	139	0	96	140	0	97	141	0	98
138	0	94	139	0	95	140	0	96	141	0	97	142	0	98
139	0	94	140	0	95	141	0	96	142	0	97	143	0	98
140	0	94	141	0	95	142	0	96	143	0	97	144	0	98
141	0	94	142	0	95	143	0	96	144	0	97	145	0	98
142	0	94	143	0	95	144	0	96	145	0	97	146	0	98
143	0	94	144	0	95	145	0	96	146	0	97	147	0	98
144	0	94	145	0	95	146	0	96	147	0	97	148	0	98
145	0	94	146	0	95	147	0	96	148	0	97	149	0	98
146	0	94	147	0	95	148	0	96	149	0	97	150	0	98
147	0	94	148	0	95	149	0	96	150	0	97	151	0	98
148	0	94	149	0	95	150	0	96	151	0	97	152	0	98
149	0	94	150	0	95	151	0	96	152	0	97	153	0	98
150	0	94	151	0	95	152	0	96	153	0	97	154	0	98
151	0	94	152	0	95	153	0	96	154	0	97	155	0	98
152	0	94	153	0	95	154	0	96	155	0	97	156	0	98
153	0	94	154	0	95	155	0	96	156	0	97	157	0	98
154	0	94	155	0	95	156	0	96	157	0	97	158	0	98
155	0	94	156	0	95	157	0	96	158	0	97	159	0	98
156	0	94	157	0	95	158	0	96	159	0	97	160	0	98
157	0	94	158	0	95	159	0	96	160	0	97	161	0	98
158	0	94	159	0	95	160	0	96	161	0	97	162	0	98
159	0	94	160	0	95	161	0	96	162	0	97	163	0	98
160	0	94	161	0	95	162	0	96	163	0	97	164	0	98
161	0	94	162	0	95	163	0	96	164	0	97	165	0	98
162	0	94	163	0	95	164	0	96	165	0	97	166	0	98
163	0	94	164	0	95	165	0	96	166	0	97	167	0	98
164	0	94	165	0	95	166	0	96	167	0	97	168	0	98
165	0	94	166	0	95	167	0	96	168	0	97	169	0	98
166	0	94	167	0	95	168	0	96	169	0	97	170	0	98
167	0	94	168	0	95	169	0	96	170	0	97	171	0	98
168	0	94	169	0	95	170	0	96	171	0	97	172	0	98
169	0	94	170	0	95	171	0	96	172	0	97	173	0	98
170	0	94	171	0	95	172	0	96	173	0	97	174	0	98
171	0	94	172	0	95	173	0	96	174	0	97	175	0	98
172	0	94	173	0	95	174	0	96	175	0	97	176	0	98
173	0	94	174	0	95	175	0	96	176	0	97	177	0	98
174	0	94	175	0	95	176	0	96	177	0	97	178	0	98
175	0	94	176	0	95	177	0	96	178	0	97	179	0	98
176	0	94	177	0	95	178	0	96	179	0	97	180	0	98
177	0	94	178	0	95	179	0	96	180	0	97	181	0	98
178	0	94	179	0	95	180	0	96	181	0	97	182	0	98
179	0	94	180	0	95	181	0	96	182	0	97	183	0	98
180	0	94	181	0	95	182	0	96	183	0	97	184	0	98
181	0	94	182	0	95	183	0	96	184	0	97	185	0	98
182	0	94	183	0	95	184	0	96	185	0	97	186	0	98
183	0	94	184	0	95	185	0	96	186	0	97	187	0	98
184	0	94	185	0	95	186	0	96	187	0	97	188	0	98
185	0	94	186	0	95	187	0	96	188	0	97	189	0	98
186	0	94	187	0	95	188	0	96	189	0	97	190	0	98
187	0	94	188	0	95	189	0	96	190	0	97	191	0	98
188	0	94	189	0	95	190	0	96	191	0	97	192	0	98
189	0	94	190	0	95	191	0	96	192	0	97	193	0	98
190	0	94	191	0	95	192	0	96	193	0	97	194	0	98
191	0	94	192	0	95	193	0	96	194	0	97	195	0	98
192	0	94	193	0	95	194	0	96	195	0	97	196	0	98
193	0	94	194	0	95	195	0	96	196	0	97	197	0	98
194	0	94	195	0	95	196	0	96	197	0	97	198	0	98
195	0	94	196	0	95	197	0	96	198	0	97	199	0	98
196	0	94	197	0	95	198	0	96	199	0	97	200	0	98
197	0	94	198	0	95	199	0	96	200	0	97	201	0	98
198	0	94	199	0	95	200	0	96	201	0	97	202	0	98
199	0	94	200	0	95	201	0	96	202	0	97	203	0	98
200	0	94	201	0	95	202	0	96	203	0	97	204	0	98
201	0	94	202	0	95	203	0	96	204	0	97	205	0	98
202	0	94	203	0	95	204	0	96	205	0	97	206	0	98
203	0	94	204	0	95	205	0	96	206	0	97	207	0	98
204	0	94	205	0	95	206	0	96	207	0	97	208	0	98
205	0	94	206	0	95	207	0	96	208	0	97	209	0	98
206	0	94	207	0	95	208	0	96	209	0	97	210	0	98
207	0	94	208	0	95	209	0	96	210	0	97	211	0	98
208	0	94	209	0	95	210	0	96	211	0	97	212	0	98
209	0	94	210	0	95	211	0	96	212	0	97	213	0	98
210	0	94	211	0	95	212	0	96	213	0	97	214	0	98
211	0	94	212	0	95	213	0	96	214	0	97	215	0	98
212	0	94	213	0	95	214	0	96	215	0	97	216	0	98
213	0	94	214	0	95	215	0	96	216	0	97			

дой строке и столбце корреляционной матрицы имеется более одного пропусков.

Кроме того, статистические модели очень сложно содержательно интерпретировать, для чего требуются большой труд квалифицированных аналитиков.

Таким образом, можно сделать *вывод*, что для моделирования такого сложного и малоисследованного объекта, каким является подсолнечник в процессе селекции, применение традиционных математических моделей является проблематичным.

По-видимому, решение поставленных задач может быть получено путем применения системно-когнитивного анализа (СК-анализ) [11; 12], – нового перспективного математического метода системного анализа, основанного на теории информации, системном анализе и когнитивном моделировании.

Метод является непараметрическим, позволяет сопоставимо обрабатывать тысячи градаций факторов и будущих состояний объекта управления при неполных (фрагментированных), зашумленных данных различной природы.

Для метода СК-анализа разработаны и методика численных расчетов, и соответствующий программный инструментарий, а также технология и методика их применения. Они прошли успешную апробацию при решении ряда задач в различных предметных областях [1–6; 8–13].

Наличие *инструментария* СК-анализа позволяет не только осуществить синтез семантической информационной модели (СИМ), но и *периодически поводить адаптацию и синтез ее новых версий*, обеспечивая тем самым отслеживание динамики предметной области и сохраняя высокую адекватность модели в изменяющихся условиях.

Таким образом, метод системно-когнитивного анализа для решения этих задач выбран в связи с тем, что он позволяет обрабатывать зашумлен-

ные фрагментарные данные больших размерностей и для него есть программный инструментарий.

В работе [11] приведен перечень этапов системно-когнитивного анализа, которые необходимо выполнить, чтобы осуществить синтез модели и исследование модели объекта управления:

I. Синтез содержательной информационной модели предметной области.

1. Когнитивная структуризация и формализация предметной области.
2. Формирование исследуемой выборки и управление ею.
3. Синтез или адаптация модели.
4. Оптимизация модели.
5. Измерение адекватности модели (внутренней и внешней, интегральной и дифференциальной валидности), ее скорости сходимости и семантической устойчивости.

II. Идентификация и прогнозирование состояния объекта управления, выработка управляющих воздействий.

1. Ввод распознаваемой выборки.
2. Пакетное распознавание.
3. Вывод результатов распознавания и их оценка.

III. Углубленный анализ содержательной информационной модели предметной области.

1. Информационный и семантический анализ классов и признаков.
2. Кластерно-конструктивный анализ классов распознавания и признаков, включая визуализацию результатов анализа в оригинальной графической форме когнитивной графики (семантические сети классов и признаков).
3. Когнитивный анализ классов и признаков (когнитивные диаграммы и диаграммы Вольфа Мерлина, нейросетевой анализ, классические и интегральные когнитивные карты).

Учитывая эти этапы СК-анализа выполним декомпозицию цели работы в ряд задач, решение которых обеспечит ее поэтапное достижение:

1. Когнитивная структуризация предметной области и формальная постановка задачи, проектирование структуры и состава исходных данных.
2. Получение исходных данных запланированного состава в той форме, в которой они накапливаются в поставляющей их организации (журналы).
3. Разработка Excel-формы для представления исходных данных.
4. Преобразование исходных данных в Excel-форму.
5. Контроль достоверности исходных данных и исправление ошибок.
6. Использование программного интерфейса для преобразования исходных данных из формы по датам в стандартную форму по фенофазам.
7. Использование программного интерфейса для преобразования исходных данных из стандартной формы по фенофазам в базы данных, используемые в инструментарии системно-когнитивного анализа (СК-анализ) – когнитивной аналитической системе "Эйдос" (система "Эйдос").
8. Синтез семантической информационной модели (СИМ).
9. Оптимизация СИМ.
10. Измерение адекватности СИМ.
11. Задача 1: *выявление* причинно-следственных зависимостей между фенотипическими признаками подсолнечника и его потребительскими свойствами;
12. Задача 2: разработка методики *прогнозирования* потребительских свойств растений подсолнечника на основе анализа их фенотипических признаков;
13. Задача 3: разработка методики *поддержки принятия решений* по отбору растений для селекции не по их потребительским свойствам, а на основе анализа фенотипических признаков.

14. Разработка принципов оценки экономической эффективности разработанных технологий при их применении:

- в научно-селекционных организациях;
- в образовательных учреждениях.

15. Исследование ограничений разработанной технологии и перспектив ее развития.

В соответствии с технологией АСК-анализа в результате проведения когнитивной структуризации предметной области были построены две справочные таблицы для кодирования исходной информации: хозяйственные свойства и внешне-наблюдаемые признаки.

Таблица 1 – КЛАССИФИКАЦИОННЫЕ ШКАЛЫ

Код	Наименование класса
1	Год
2	Урожайность (г.)
3	Оборы масла (г.)
4	Средняя масличность (%)
5	Натура (г/л)

Таблица 2 – ОПИСАТЕЛЬНЫЕ ШКАЛЫ (ФАКТОРЫ)

Код	Наименование класса
1	Вес семян с 1 корзины (г.)
2	Время посева
3	Высота растения (см.)
4	Обпиственность (г.)
5	Диаметр корзины (см.)

Затем была выполнена формализация предметной области, т.е. разработаны справочники (табл. 3 и 4).

Таблица 3 – КЛАССИФИКАЦИОННЫЕ ШКАЛЫ И ГРАДАЦИИ (КЛАССЫ)

Код	Наименование класса
1	Год
1	1994
2	1995
3	1996
4	1997
5	1998
6	1999
7	2000
8	2001
9	2002
10	2003
2	Урожайность (г.)
11	менее 400
12	[400, 500)
13	[500, 600)
14	[600, 700)
15	[700, 800)
16	800 и более
3	Сборы масла (г.)
17	менее 200
18	[200, 250)
19	[250, 300)
20	[300, 350)
21	[350, 400)
22	400 и более
4	Средняя масличность (%)
23	менее 50
24	[50, 51)
25	51, 52
26	52, 53
27	53, 54
28	54 и более
5	Натура (г/л)
29	менее 420
30	420, 430
31	430, 440
32	440, 450
33	450, 460
34	460, 470
35	470 и более

Таблица 4 – ОПИСАТЕЛЬНЫЕ ШКАЛЫ И ГРАДАЦИИ (ФАКТОРЫ)

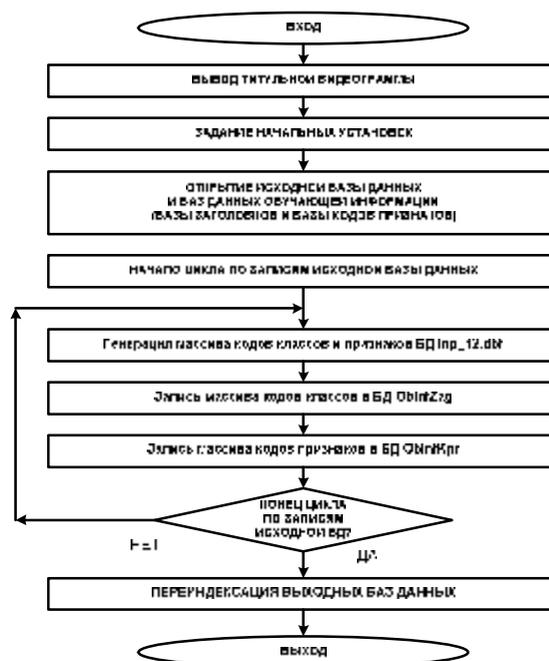
Код призна	Наименования шкал и градаций признаков
[1]	ВЕС СЕМЯН С 1 КОРЗИНЫ (Г)
1	менее 60
2	[60, 70)
3	[70, 80)
4	[80, 90)
5	[90, 100)
6	100 и более
[2]	ВРЕМЯ ПОСЕВА
7	ПОЗДНЕЗИМНИЙ (ДЕКАБРЬ)
8	РАННЕВЕСЕННИЙ (ФЕВРАЛЬ)
9	ОПТИМАЛЬНЫЙ
[3]	ВЫСОТА РАСТЕНИЯ (см.)
10	менее 50
11	[50, 80)
12	[80, 110)
13	[110, 140)
14	[140, 170)
15	[170, 200)
16	200 и более
[4]	ОБЛИСТВЕННОСТЬ (шт.)
17	менее 30
18	30
19	31
20	32
21	33
22	34
23	35
24	35 и более
[5]	ДИАМЕТР КОРЗИНЫ (см.)
25	менее 19
26	19
27	20
28	21
29	22
30	23
31	24
32	24 и более

С их использованием этих справочников произведено кодирование исходных данных. Результаты кодирования исходных данных из журналов с помощью разработанных классификационных и описательных шкал представлены EXCEL-таблице, фрагмент которой представлен в таблице 6.

Таблица 6 – ФРАГМЕНТ EXCEL-ТАБЛИЦЫ ДЛЯ ВВОДА ИСХОДНЫХ ДАННЫХ

№ при- мера	Год, № деловки	Коды классов					Коды значимых факторов				
		03	04	05	06	07	08	09	10	11	12
1	1994-1521	1	14	20	20	32	7	7	16	18	28
2	1994-1522	1	14	21	28	35	4	7	16	17	28
3	1994-1523	1	15	22	28	34	4	7	15	22	28
4	1994-1527	1	15	21	28	32	5	7	16	19	27
5	1994-1528	1	15	21	20	33	5	7	10	20	28
6	1994-1529	1	14	20	23	35	5	7	15	20	27
7	1994-1535	1	12	18	27	35	7	7	15	17	25
8	1994-1537	1	13	20	25	32	4	7	15	19	27
9	1994-1541	1	16	22	26	30	3	7	15	19	26
10	1994-1545	1	14	20	27	33	3	7	15	17	20

Разработан программный интерфейс, обеспечивающий автоматический ввод обучающей выборки из Excel-файла с входной информацией в соответствующие базы данных системы "Эйдос", которая представляет собой инструментарий АСК-анализа.



Дальнейшие этапы выполнялись с помощью системы «Эйдос» в подсистеме синтеза модели. При этом на основе обучающей выборки были автоматически сформированы матрица абсолютных частот и матрица информативностей.

Атрибуты	Классы - будущие состояния объекта управления					Сумма
	Целевые состояния		Индикаторные состояния			
	***	j	***	i	***	
Факторы, характеризующие текущее и прошлые состояния объекта управления, а также его рефлексивность	***					
	i	N_{ij}		N_i		$N = \sum_{i=1}^w N_i$

Управляющие факторы системы управления	i	N_{ij}		N_{ij}		$N_i = \sum_{j=1}^w N_{ij}$

Факторы, характеризующие прошлые, текущее и прогнозируемые состояния окружающей среды	k	N_{kj}		N_k		$N_k = \sum_{j=1}^w N_{kj}$

Сумма		$N_j = \sum_{i=1}^w N_{ij}$		$N_i = \sum_{j=1}^w N_{ij}$		$N = \sum_{i=1}^w \sum_{j=1}^w N_{ij}$

где:

N_{ij} - количество встреч i-го признака у объектов j-го класса по данным обучающей выборки

Атрибут	Классы - будущие состояния объекта управления			Средняя детерминированность исходных факторов
	Целевые состояния		Индикаторные состояния	
	***	j	***	i
Факторы, характеризующие текущее и прошлые состояния объекта управления, а также его рефлексивность	***			
	$I_{i*} = \Psi \cdot \text{Log}_2 \frac{N_{i*} \cdot \sum_{j=1}^w \sum_{k=1}^m N_{kj}}{\sum_{j=1}^w N_{ij} \cdot \sum_{k=1}^m N_{kj}}$			$I_{i*} = \Psi \cdot \text{Log}_2 \frac{N_{i*} \cdot \sum_{j=1}^w \sum_{k=1}^m N_{kj}}{\sum_{j=1}^w N_{ij} \cdot \sum_{k=1}^m N_{kj}}$

Управляющие факторы системы управления				
	$I_{i*} = \Psi \cdot \text{Log}_2 \frac{N_{i*} \cdot \sum_{j=1}^w \sum_{k=1}^m N_{kj}}{\sum_{j=1}^w N_{ij} \cdot \sum_{k=1}^m N_{kj}}$			$I_{i*} = \Psi \cdot \text{Log}_2 \frac{N_{i*} \cdot \sum_{j=1}^w \sum_{k=1}^m N_{kj}}{\sum_{j=1}^w N_{ij} \cdot \sum_{k=1}^m N_{kj}}$

Факторы, характеризующие прошлые, текущее и прогнозируемые состояния окружающей среды	k			
	$I_{i*} = \Psi \cdot \text{Log}_2 \frac{N_{i*} \cdot \sum_{j=1}^w \sum_{k=1}^m N_{kj}}{\sum_{j=1}^w N_{ij} \cdot \sum_{k=1}^m N_{kj}}$			$I_{i*} = \Psi \cdot \text{Log}_2 \frac{N_{i*} \cdot \sum_{j=1}^w \sum_{k=1}^m N_{kj}}{\sum_{j=1}^w N_{ij} \cdot \sum_{k=1}^m N_{kj}}$

Средняя детерминированность будущих состояний ГОУ				
	$\sigma_i = \sqrt{\frac{1}{M-1} \sum_{j=1}^M (I_{ij} - \bar{I}_i)^2}$			$\sigma_i = \sqrt{\frac{1}{M-1} \sum_{j=1}^M (I_{ij} - \bar{I}_i)^2}$

$\bar{I}_i = \frac{1}{M} \sum_{j=1}^M I_{ij}$ - среднее значение информации в векторе I_{i*} ; M - количество векторов; $\bar{I} = \frac{1}{W} \sum_{i=1}^w \bar{I}_i$ - среднее значение информации в векторе факторов; W - число состояний классов (будущих состояний ГОУ); $I = \frac{1}{W \cdot M} \sum_{i=1}^w \sum_{j=1}^M I_{ij}$ - средняя информативность системы по признаку информативности; $\sigma = \frac{\text{Log}_2 \sum_{i=1}^w C_i}{\text{Log}_2 W}$ - коэффициент акредитивности Керри; $\psi = \frac{\text{Log}_2 W^2}{\text{Log}_2 N}$ - коэффициент акредитивности Карнеги; H - мера уровня достоверности прогноза объекта в рамках ГОУ

После синтеза семантической информационной модели, была изменена её адекватность, средневзвешенная достоверность прогнозирования, которая составила 66,42 %, что достаточно для поставленных целей и принятия решений.

ИЗМЕРЕНИЕ АДЕКВАТНОСТИ СЕМАНТИЧЕСКОЙ ИНФОРМАЦИОННОЙ МОДЕЛИ

Пакет физических: 1000 логических (всего/факт): 5000/ 5000
 Верная идентификация: 3321 Ошибочная идентификация: 1679
 Верная идентификация: 66.42% Ошибочная идентификация: 33.58%
 Минимальный уровень сходства: 0.0 Максимальное кол-во классов: 99999

г.Краснодар

N п/п	Код класса	Полное название класса	Всего логич. пакет	ИДЕНТИФИЦИР		Неидентифицир		ИДЕНТИФИЦИРОВ		Неидентифицир	
				ВЕРНО	Ошиб.	Верно	Ошиб.	ВЕРНО%	Ошиб.%	Верно%	Ошиб.%
1	1	ГОД - 1994	100	87	401	409	13	87.00	0.80	9.57	13.00
2	7	ГОД - 2000	100	70	451	449	22	70.00	9.20	9.16	22.00
3	2	ГОД - 1995	100	77	473	427	23	77.00	9.65	8.71	23.00
4	5	ГОД - 1998	100	74	461	439	26	74.00	9.41	8.96	26.00
5	6	ГОД - 1999	100	73	454	446	27	73.00	9.27	9.10	27.00
6	4	ГОД - 1997	100	72	497	403	28	72.00	10.14	8.22	28.00
7	25	СРЕДНЯЯ ИМПЛИЦИТНОСТЬ (51, 52)	163	117	459	378	46	71.78	9.49	7.81	28.22
8	16	ЧРЯЖАЮЩЕСТВО - 800 и более	159	110	440	398	49	69.18	9.15	0.22	30.82
9	17	СБОРЫ ИСПИ - менее 200	160	116	422	410	52	69.05	0.73	0.49	30.95
10	8	ГОД - 1996	100	69	427	473	31	69.00	8.71	9.65	31.00
11	33	НАТУРА - (450, 460)	152	104	405	443	48	68.42	8.35	9.14	31.58
12	30	НАТУРА - (420, 430)	136	93	414	450	43	60.00	0.51	9.25	31.62
13	22	СБОРЫ ИСПИ - 400 и более	158	108	426	416	50	68.35	8.80	8.59	31.65
14	8	ГОД - 2001	100	68	424	476	32	68.00	8.65	9.71	32.00
15	31	НАТУРА - (430, 440)	159	107	450	391	52	67.00	9.00	0.00	32.70
16	27	СРЕДНЯЯ ИМПЛИЦИТНОСТЬ (53, 54)	166	111	401	433	55	66.87	8.30	8.96	33.13
17	12	УРОЖАНОСТЬ (400, 500)	165	110	398	437	55	66.67	8.23	9.04	33.33
18	9	ГОД - 2002	100	66	409	461	34	66.00	0.96	9.41	36.00
19	10	УРОЖАНОСТЬ - (500, 600)	164	100	452	304	56	65.85	9.05	7.94	36.15
20	28	СРЕДНЯЯ ИМПЛИЦИТНОСТЬ 54 и более	184	120	410	406	64	65.22	8.51	8.43	36.78
21	19	СБОРЫ ИСПИ - (250, 300)	158	103	410	432	55	65.19	8.47	8.92	36.81
22	21	СБОРЫ ИСПИ - (350, 400)	154	100	382	464	54	64.94	7.80	9.57	35.06
23	35	НАТУРА (470 и более	136	88	459	405	48	64.71	9.44	8.33	35.29
24	32	НАТУРА (440, 450)	132	85	393	475	47	64.39	8.07	9.76	35.61
25	24	СРЕДНЯЯ ИМПЛИЦИТНОСТЬ - (50, 51)	160	100	404	428	60	64.29	0.00	0.80	35.71
26	11	УРОЖАНОСТЬ - менее 400	144	92	410	446	52	63.89	0.44	9.10	36.11
27	15	УРОЖАНОСТЬ (700, 800)	173	110	389	438	63	63.58	8.06	9.07	36.42
28	18	СБОРЫ ИСПИ - (200, 250)	171	108	421	408	63	63.16	8.72	8.45	36.84
29	20	СРЕДНЯЯ ИМПЛИЦИТНОСТЬ - менее 50	170	109	392	435	64	63.01	0.12	9.01	36.99
30	10	ГОД - 2003	100	63	439	461	37	63.00	8.96	9.41	37.00
31	14	УРОЖАНОСТЬ (600, 700)	195	120	377	428	75	61.54	7.85	8.21	38.46
32	29	НАТУРА - менее 420	150	91	400	447	59	60.67	0.01	9.22	39.33
33	26	СРЕДНЯЯ ИМПЛИЦИТНОСТЬ - (52, 53)	146	87	390	461	59	59.59	0.10	9.50	40.41
34	20	СБОРЫ ИСПИ (300, 350)	191	112	362	447	79	58.64	7.53	9.30	41.36
35	34	НАТУРА - (460, 470)	135	77	405	460	58	57.04	8.32	9.46	42.96

Универсальная когнитивная аналитическая система

НИИ "Эйдос"

Получена матрица информативности с конкретными значениями силы и направления причинно-следственных связей между фенотипическими и хозяйственными свойствами подсолнечника.

На основе полученной матрицы информативности для каждого конкретного растения по его внешним признакам можно прогнозировать его хозяйственные свойства. Результаты прогнозирования выводятся системой "Эйдос" в специальной форме, каждая строка которой соответствует классу, с которым данный объект имеет наибольшее сходство.

Задача принятия решений является обратной по отношению к задаче прогнозирования. Если при прогнозировании мы по фенотипическим признакам определяем возможные хозяйственные свойства, то при принятии решений, наоборот, по заданным хозяйственным свойствам определяем

какими фенотипическим признаками должны обладать растения, имеющие эти свойства.

Получен основной результат, позволяющий принимать решения по отбору растений для следующих поколений по их фенотипическим признакам, отражающий какие внешние признаки характерны для тех или иных хозяйственных свойств.

Рассмотрим эти результаты подробнее.

Задача 1: выявление причинно-следственных зависимостей между фенотипическими признаками подсолнечника и его потребительскими свойствами. Мерой причинно-следственных зависимостей в АСК-анализе является количество информации, которое содержится в факте действия определенного значения фактора о том, что растение подсолнечника будет иметь определенное потребительское свойство. Значение информативности может быть по модулю различной величины и положительным и отрицательным по знаку, что означает, соответственно, величину и направление влияния данного значения фактора на данное потребительское свойство.

В полном виде все зависимости содержатся в матрице информативностей. Вертикальная шапка данной матрицы содержит градации описательных, в горизонтальная шапка – градации классификационных шкал (табл. 7).

Таблица 7 – Матрица информативности (фрагмент)

	1	2	3	4	5	6	7	8	9	10	11	12
1	-0,21	0,02	-0,03	0,02	0,05	0,10	-0,00	-0,03	-0,00	0,02	-0,06	0,14
2	-0,24	-0,03	-0,07	-0,07	0,11	0,00	0,14	0,06	-0,03	0,03	0,06	-0,03
3	0,16	0,01	-0,08	0,01	-0,08	-0,08	-0,01	0,04	0,01	-0,05	0,03	-0,03
4	0,09	0,06	0,06	-0,04	-0,08	-0,08	-0,04	-0,04	0,06	-0,04	0,03	0,02
5	0,04	-0,15	0,01	0,09	-0,08	-0,01	-0,11	0,07	0,01	0,07	0,05	-0,05
6	0,06	0,06	0,09	-0,06	0,03	0,03	0,00	-0,14	-0,06	-0,06	-0,15	-0,09
7	-0,04	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,04	0,07

8	-0,23	0,02	0,02	0,02	0,02	0,02	0,02	0,02	0,02	0,02	0,11	-0,05
9	0,17	-0,02	-0,02	-0,02	-0,02	-0,02	-0,02	-0,02	-0,02	-0,02	-0,18	-0,02
10	-0,47	-0,12	0,18	-0,17	0,18	-0,23	0,08	0,15	-0,03	0,08	0,12	0,20
11	-0,31	-0,10	0,16	0,01	0,01	0,13	0,04	0,04	0,04	-0,20	0,18	-0,15
12	-0,41	-0,10	-0,06	0,12	0,12	-0,06	0,02	-0,02	0,15	0,05	-0,07	-0,05

Из-за большой размерности данной матрицы полностью ее привести в данной работе не представляется возможным.

Задача 2: разработка методики прогнозирования потребительских свойств растений подсолнечника на основе анализа их фенотипических признаков. Данная задача решается на основе задачи 1. Суть методики прогнозирования состоит в следующем. Из матрицы информативностей известно, какое количество информации содержится в определенном фенотипическом признаке конкретного растения подсолнечника о том, что данное растение будет обладать каждым из исследованных в модели потребительских свойств. Если о растении известно, что оно обладает определенным набором фенотипических признаков, то естественно считать, что оно будет обладать теми потребительскими свойствами, о которых в данном наборе содержится наибольшее количество информации. Данный интегральный критерий предложен и обоснован в [11].

Исходные данные для прогнозирования вводятся в 1-м режиме 4-й подсистемы системы «Эйдос» в форме распознаваемой выборки, само прогнозирование осуществляется во 2-м режиме этой же подсистемы, а результаты выводятся в 3-м режиме в двух разрезах:

- один объект – много классов;
- один класс – много объектов.

Результаты прогнозирования выводятся системой в обобщенной форме, каждая строка которой соответствует классу, с которым данный объект имеет наиболее сходство, и в детализированной форме карточек

прогнозирования (распознавания). Примеры этих карточек приведены на рисунке 1.

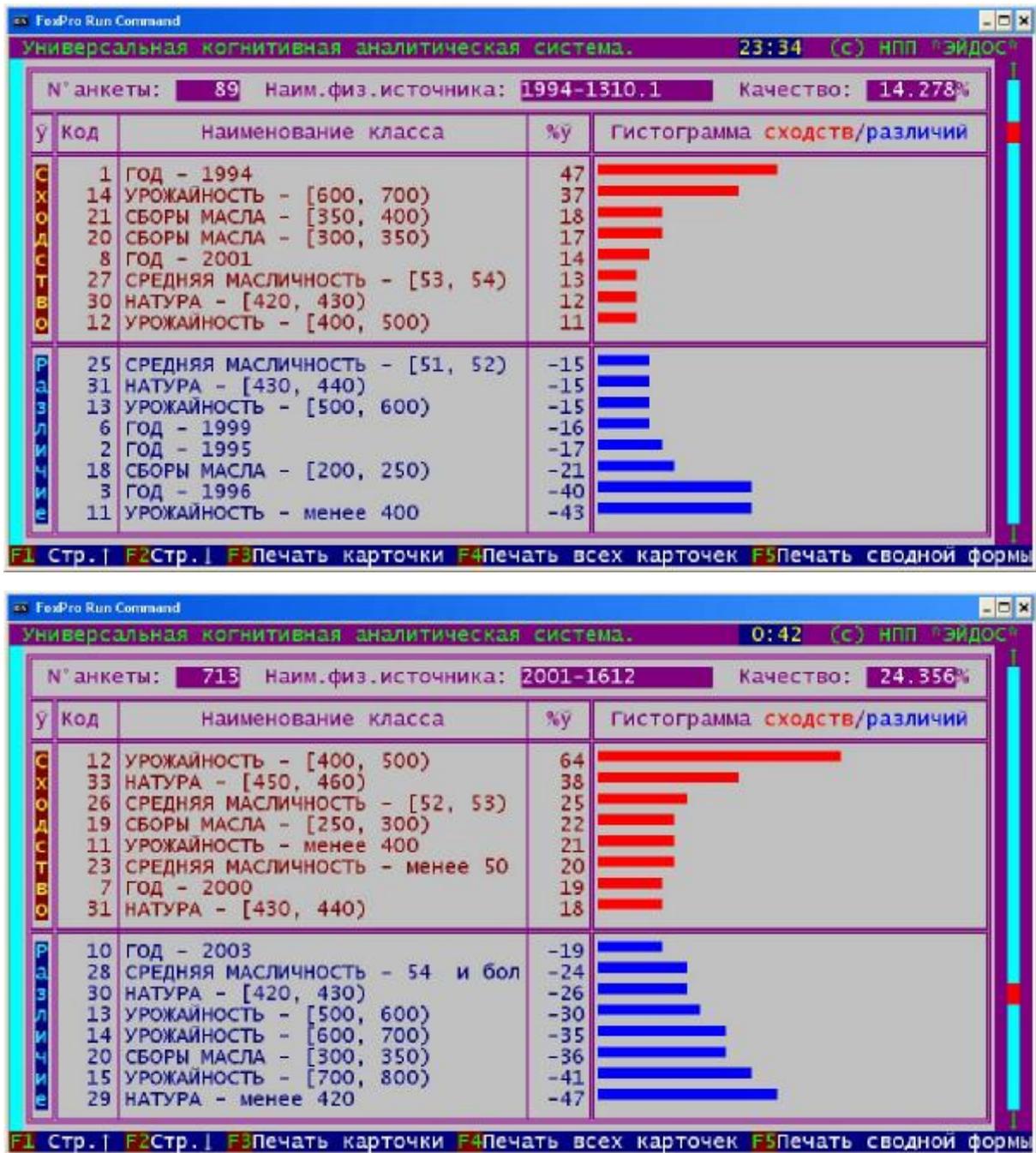


Рисунок 1 – Примеры карточки прогнозирования (экранные формы)

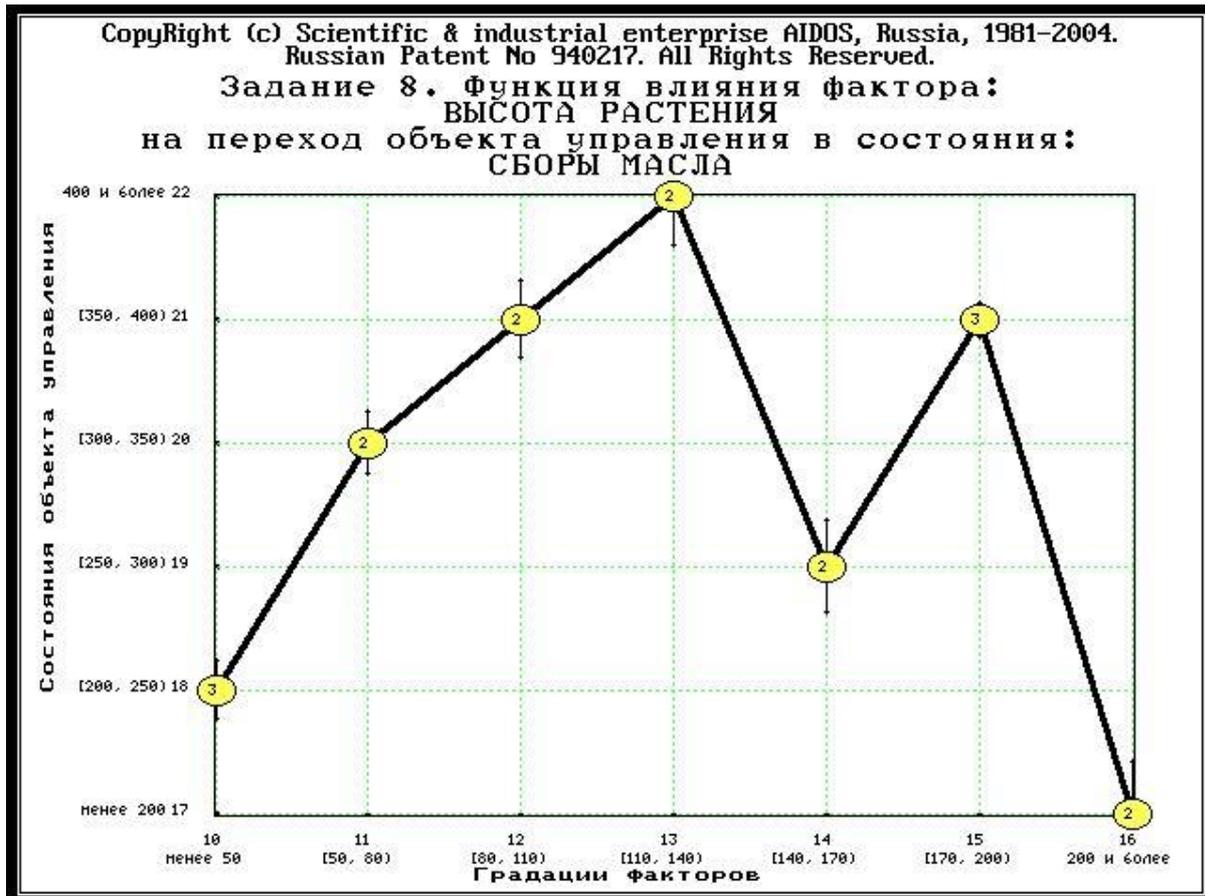
Карточка разделена на две части. В верхней части приведены классы, с которыми данный объект имеет наивысшее сходство в порядке его убывания, а в нижней – классы, от которых данный объект максимально отли-

чается. Задача идентификации может решаться в поле с использованием ноутбука, на котором установлена система «Эйдос».

Задача 3: разработка методики поддержки принятия решений по отбору растений для селекции не по их потребительским свойствам, а на основе анализа фенотипических признаков. Задача принятия решений является обратной по отношению к задаче прогнозирования. Если при прогнозировании мы по фенотипическим признакам определяем возможные потребительские свойства, то при принятии решений наоборот, по заданным потребительским свойствам определяем какими фенотипическими признаками должны обладать растения, имеющие эти свойства. В системе «Эйдос» эта задача решается в 1-м режиме 5-й подсистемы, который позволяет генерировать и отображать так называемые «Информационные портреты классов». Эти информационные портреты показывают систему детерминации будущих состояний объекта управления, в нашем случае – потребительских свойств подсолнечника.

Они могут непосредственно использоваться при отборе растений подсолнечника для формирования следующего поколения при селекции (табл. 9).

Таблица 9 – Система детерминации хозяйственно-значимых свойств подсолнечника фенотипическими признаками



**Рисунок 2 – Взаимосвязь между высотой растения
и сбором с него масла**

Из этого графика очевидно, что оптимальной высотой растения является 110–140 см, а не 200, как считалось ранее. Объяснение причин этого явления сходно с содержательной интерпретацией взаимосвязи между количеством листьев на растении и масличностью собранных с него семян, представленная на рисунке 3.

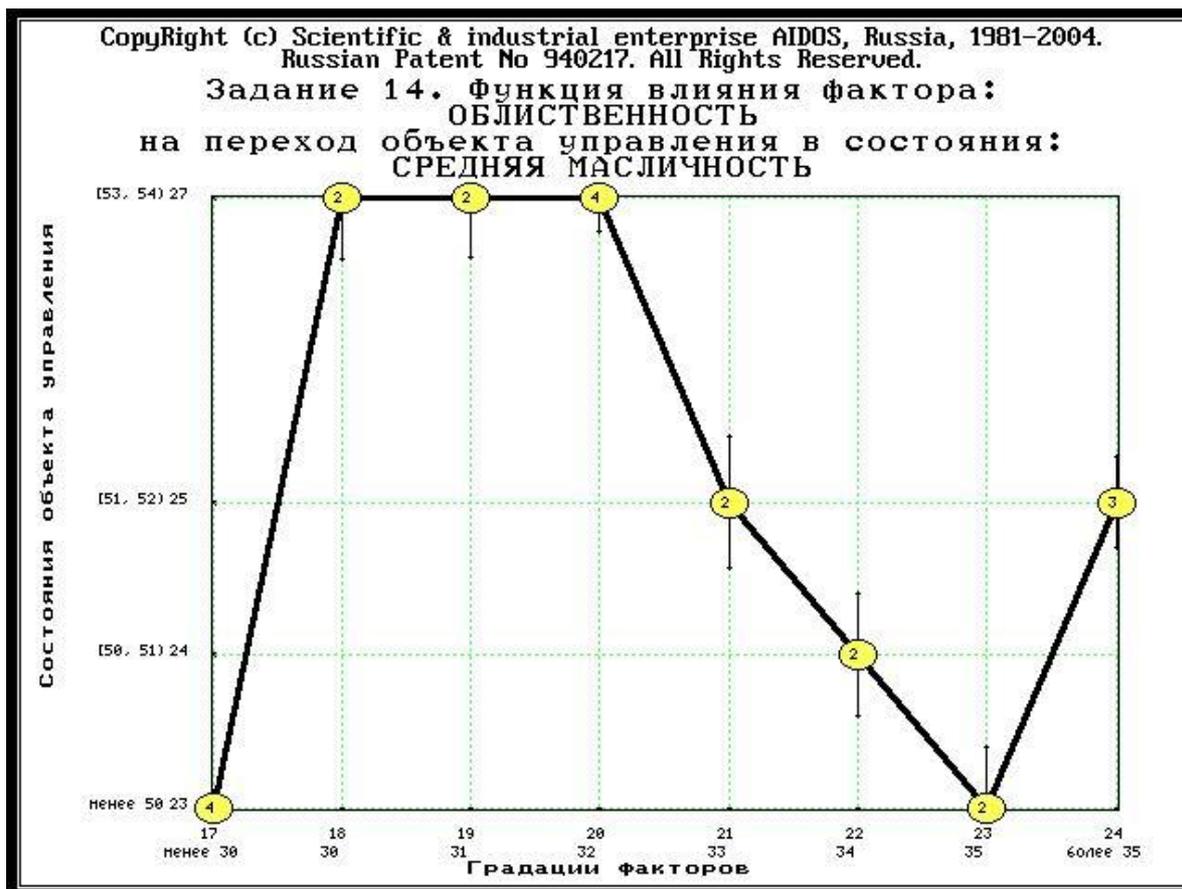


Рисунок 3 – Взаимосвязь между количеством листьев на растении и масличностью собранных с него семян

При малом количестве листьев растение не имеет необходимых ресурсов для генерации масла в семенах, при количестве листьев 30–32 наблюдается максимальная масличность, при увеличении количества листьев масличность закономерно падает, что по-видимому можно объяснить тем, что ресурсы растения переориентируются с формирования регенеративных качеств растения на увеличение его вегетативной массы. Этот эффект, наблюдаемый у подсолнечника, полностью аналогичен известному «бройлерному эффекту», известному в птицеводстве и впервые был обнаружен методом СК-анализа эмпирических данных в 1993 году О. А. Засухиной и Е. В. Луценко при исследовании интенсивных технологий возделывания зерновых колосовых [11].

Все функции влияния, отражающие взаимосвязи между всеми описательными и классификационными шкалами и градациями (во всех сочетаниях), в данной статье привести нет возможности. Необходимо отметить, что их содержательная интерпретация является делом ученого-селекционера.

Кластерно-конструктивный и системно-когнитивный анализ результатов выращивания и факторов обеспечивается режимами 5-й подсистемы "Типология" системы «Эйдос» (рис. 4).

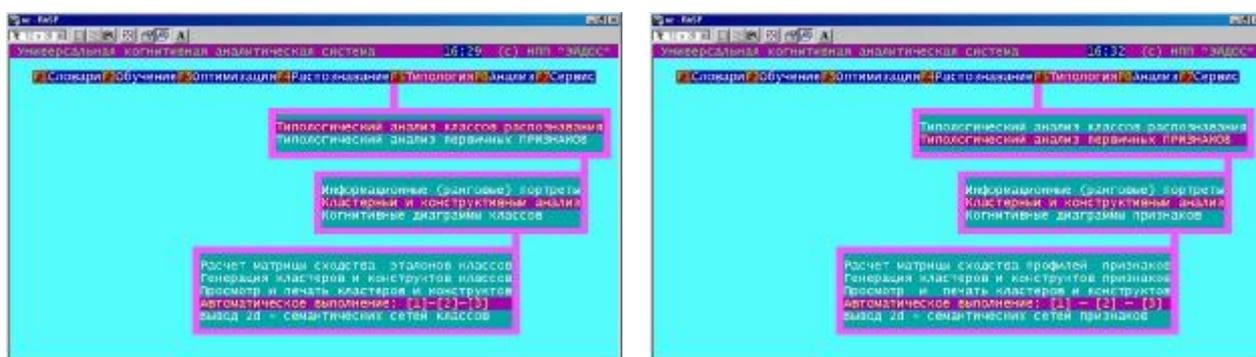


Рисунок 4 – Режимы подсистемы типологического анализа классов и факторов

Кластерный анализ классов показывает, какие качественные и количественные результаты выращивания различных сортов детерминируются (вызываются) сходными системами факторов, и могут быть получены одновременно, а какие противоположными, несовместимыми и одновременно недостижимыми. В графической форме любые заданные фрагменты матрицы сходства могут отображаются в виде семантических сетей (рис. 5, 6 и 7).

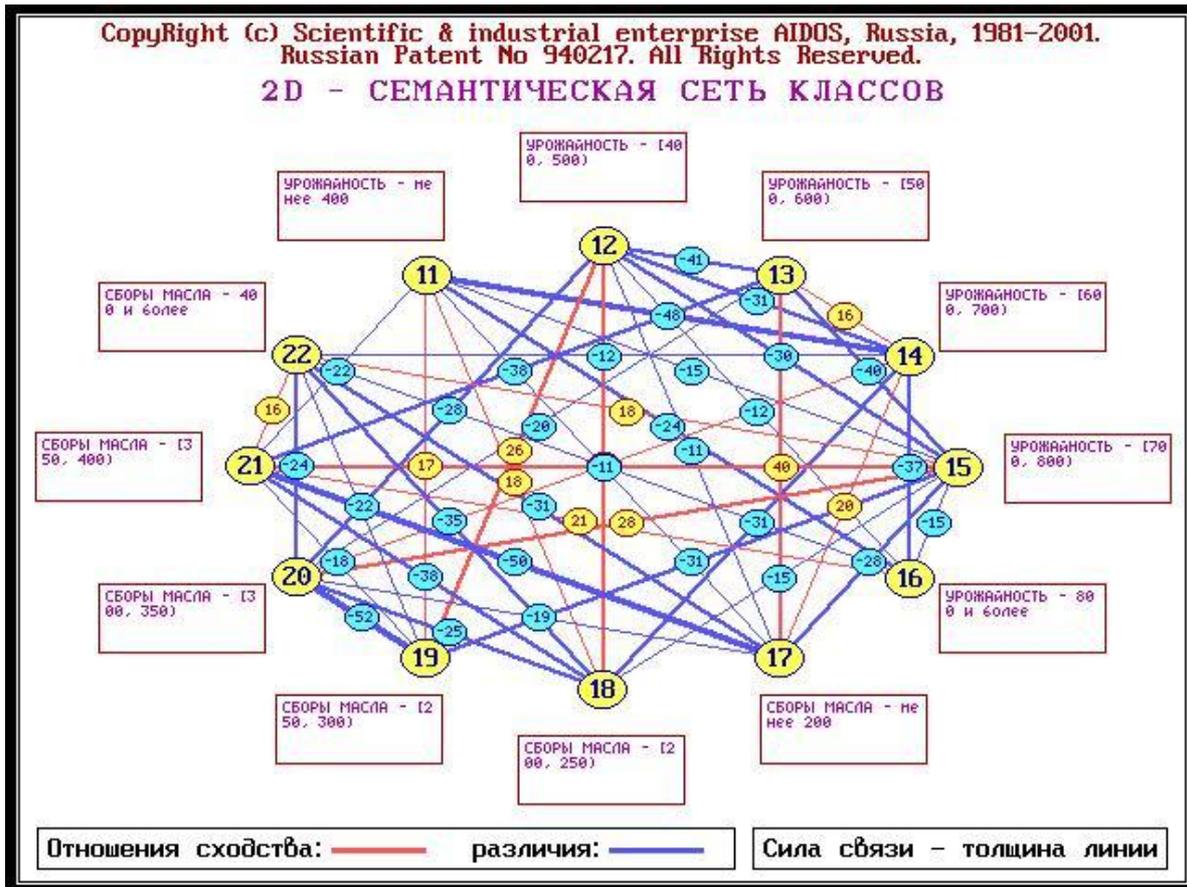


Рисунок 5 – Семантическая сеть классов

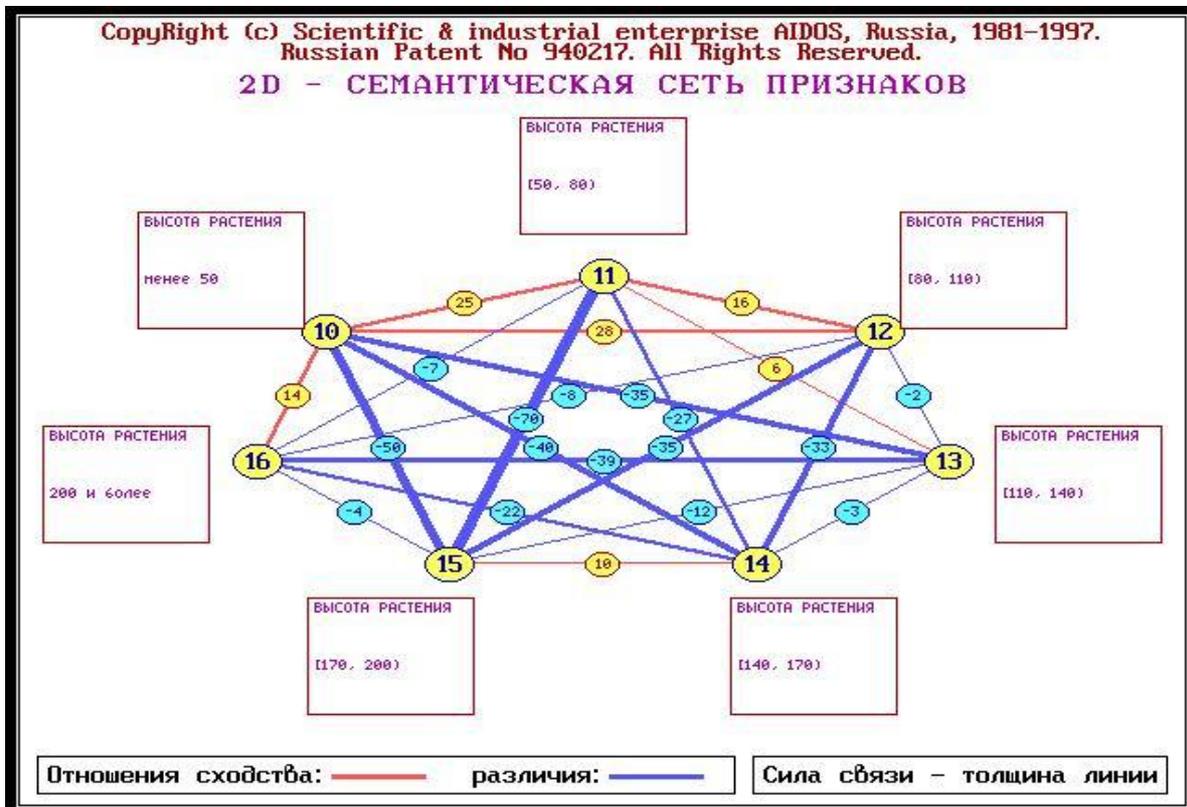


Рисунок 6 – Семантическая сеть фенотипических признаков:
«Высота растения»

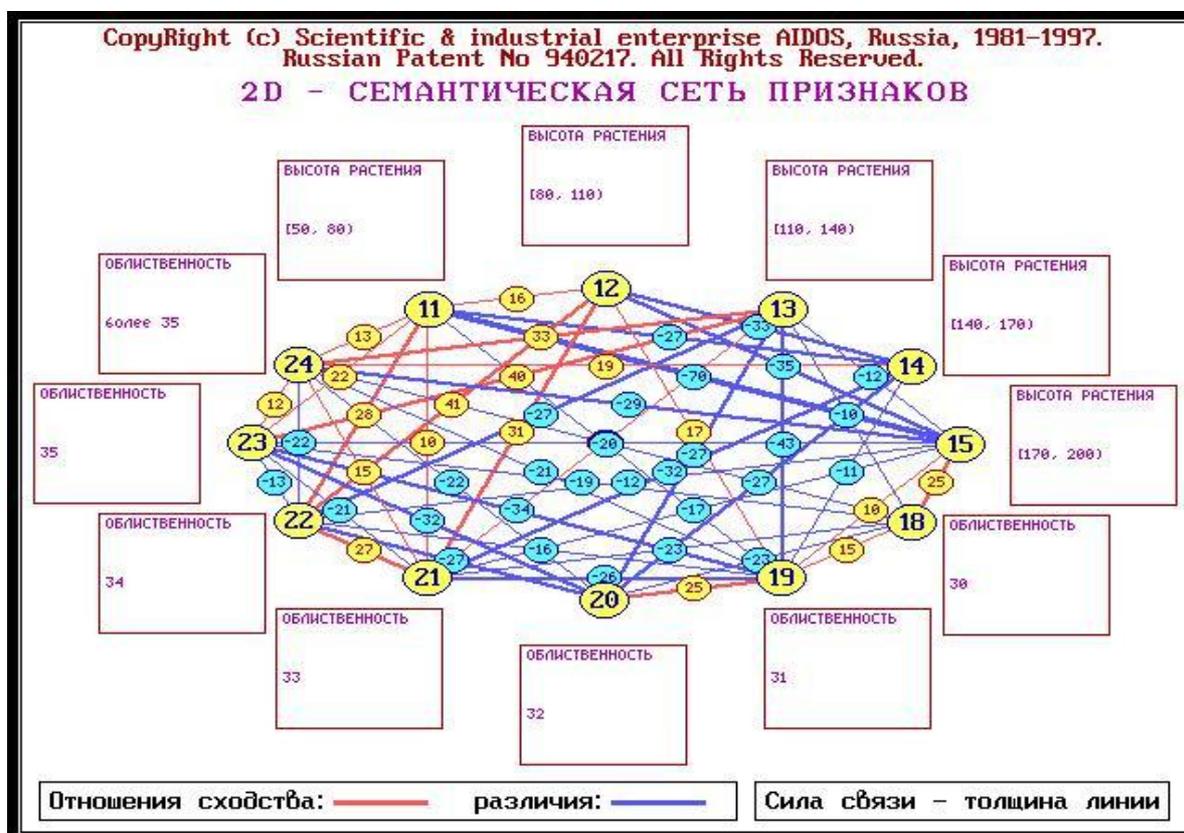


Рисунок 7 – Семантическая сеть фенотипических признаков:
«Облиственность» и «Высота растения»

Итак, в результате проведенной работы осуществлены когнитивная структуризация и формализация предметной области, спроектирована структура исходных данных; получены исходные данные в форме агрономических журналов; разработана Excel-форма для ввода исходных данных; исходные данные введены в Excel-форму; выполнены контроль достоверности исходных данных и исправление ошибок ввода; разработан программный интерфейс, с помощью которого исходные данные импортированы в базы данных, используемые в инструментарии системно-когнитивного анализа – когнитивной аналитической системе "Эйдос"; в

этой системе осуществлен синтез семантической информационной модели (СИМ); выполнена оптимизация СИМ; измерена адекватность СИМ.

Получены следующие основные результаты:

- решена задача 1: *выявлены* причинно-следственные зависимости между фенотипическими признаками подсолнечника и его хозяйственными свойствами;

- решена задача 2: разработана методика *прогнозирования* хозяйственных свойств растений подсолнечника на основе анализа фенотипических признаков;

- решена задача 3: Разработана методика *поддержки принятия решений* по отбору растений для селекции не по их хозяйственным (потребительским) свойствам, а на основе анализа фенотипических признаков;

- разработаны принципы оценки экономической эффективности разработанных методов компьютерной селекции при их применении: в научно-селекционных и в образовательных учреждениях;

- исследованы ограничения разработанной технологии и обоснованы перспективы ее развития.

Решение данной проблемы позволит избежать инструментального изучения хозяйственных свойств и генотипа при отборе растений для следующих селекционных поколений, что позволяет получить существенную экономию различных ресурсов и является актуальным как для науки, так и для практики селекции и может быть использовано в селекционных организациях, а также в учебном процессе КубГАУ при проведении лабораторных работ по дисциплине: "Интеллектуальные информационные системы", изучаемой на 5-м курсе специальности 351400 – Прикладная информатика.

Список литературы

1. Драгавцева И. А., Луценко Е. В., Лопатина Л. М. Автоматизированная система мониторинга, анализа и прогнозирования развития сельхозкультур "ПРОГНОЗ-АГРО". Пат. № 2003610433 РФ. Заяв. № 2002611927 РФ. Опубл. от 18.02.2003.
2. Драгавцева И. А., Луценко Е. В., Лопатина Л. М. База данных автоматизированной системы мониторинга, анализа и прогнозирования развития сельхозкультур "ПРОГНОЗ-АГРО". Пат. № 2003620035 РФ. Заяв. № 2002620178 РФ. Опубл. от 20.02.03.
3. Драгавцева, И. А. Автоматизация системного анализа продуктивности плодовых культур Юга России / И. А. Драгавцева, Е. В. Луценко, Л. М. Лопатина // Научные труды Российской академии сельскохозяйственных наук. – 2002. – С.11–14.
4. Драгавцева, И. А. Применение системного анализа для прогнозирования успешности выращивания сельскохозяйственных культур (на примере плодовых) / И. А. Драгавцева, Е. В. Луценко, Л. М. Лопатина, Н. Е. Луценко // в сб. "Формы и методы повышения эффективности координации исследований для ускорения процесса передачи реальному сектору экономики завершенных разработок". – Краснодар : СКЗНИИСиВ, 2002. – С. 62–67.
5. Драгавцева, И. А. Применение автоматизированного системного анализа для прогноза продуктивности плодовых культур на Юге страны / И. А. Драгавцева, Е. В. Луценко, Л. М. Лопатина, Н. Е. Луценко // Материалы Всероссийской научно-практической конференции. – Краснодар : СКЗНИИСиВ, 2002. – С. 8–11.
6. Егоров, Е. А. Интенсивные технологии возделывания плодовых культур / Е. А. Егоров [и др.] : Монография (научное издание). – Краснодар : СКЗНИИСиВ, 2004. – 394 с.
7. Калайджан, А. А. Индуцированные макромутанты подсолнечника в биологизации защиты растений / А. А. Калайджан, Н. Г. Малюга // Трансгенные растения – новое направление в биологической защите растений. – Краснодар, 2003. – С. 233–238.
8. Калустов, А. А. Применение автоматизированного системно-когнитивного анализа для совершенствования методов компьютерной селекции подсолнечника / А. А. Калустов, Е. В. Луценко // Научный журнал КубГАУ [Электронный ресурс]. – Краснодар : КубГАУ, 2005. – № 02(10). – Режим доступа: <http://ej.kubagro.ru/2005/02/10/p10.asp>

9. Лопатина, Л. М. Создание автоматизированной системы мониторинга, анализа, прогноза и управления продуктивностью сельскохозяйственных культур / Л. М. Лопатина, Е. В. Луценко // Научный журнал КубГАУ [Электронный ресурс]. – Краснодар : КубГАУ, 2003. – № 02(2). – Режим доступа: <http://ej.kubagro.ru/2003/02/07/p07.asp>.
10. Лопатина, Л. М. Концептуальная постановка задачи: "Прогнозирование количественных и качественных результатов выращивания заданной культуры в заданной точке" / Л. М. Лопатина, Е. В. Луценко, И. А. Драгавцева // Научный журнал КубГАУ [Электронный ресурс]. – Краснодар : КубГАУ, 2004. – № 05(7). – Режим доступа: <http://ej.kubagro.ru/2004/05/08/p08.asp>.
11. Луценко, Е. В. Автоматизированный системно-когнитивный анализ в управлении активными объектами (системная теория информации и ее применение в исследовании экономических, социально-психологических, технологических и организационно-технических систем) : Монография (научное издание). – Краснодар : КубГАУ, 2002. – 605 с.
12. Луценко, Е. В. Интеллектуальные информационные системы : учебное пособие для студентов специальности: 351400 "Прикладная информатика (по отраслям)". – Краснодар : КубГАУ, 2004. – 633 с.
13. Максимов, А. М. Прогнозирование продуктивности и качества плодовых культур на основе применения систем искусственного интеллекта (по данным ЗАО Агрофирмы "Сад Гигант" Славянского района Краснодарского края) : Дипломная работа под научн. рук. проф. Луценко Е. В. – Краснодар : КубГАУ, 2004. – 100 с.