УДК 004.8

UDC 004.8

05.00.00. Технические науки

Technical sciences

АСК-АНАЛИЗ ЗАВИСИМОСТИ ОПЛАТЫ СОТРУДНИКОВ АПК ОТ ИХ ХАРАКТЕРИСТИК

ASC-ANALYSIS OF THE DEPENDENCE OF PAYMENTS TO EMPLOYEES OF AIC FROM THEIR CHARACTERISTICS

Луценко Евгений Вениаминович
д.э.н., к.т.н., профессор
Scopus Author ID: 57191193316
РИНЦ SPIN-код: 9523-7101
prof.lutsenko@gmail.com

Lutsenko Eugeny Veniaminovich
Dr.Sci.Econ., Cand.Tech.Sci., professor
Scopus Author ID: 57191193316
RSCI SPIN-code: 9523-7101
prof.lutsenko@gmail.com

Мартиросов Владимир Гарриевич
студент группы: ПИ1301
Кубанский государственный аграрный университет имени И.Т.Трубилина, Краснодар, Россия

Martirosov Vladimir Garrievich
Student groups: PI1301
Kuban State Agrarian University, Krasnodar, Russia

Создание систем искусственного интеллекта является одним из важных и перспективных направлений развития современных информационных технологий. Так как существует множество альтернатив систем искусственного интеллекта, то возникает необходимость оценки качества математических моделей этих систем. В данной работе рассмотрено решение задачи идентификации классов уровней оплаты сотрудников фирмы по их характеристикам. Для достижения поставленной цели необходимы свободный доступ к тестовым исходным данным и методика, которая поможет преобразовать эти данные в форму, которая необходима для работы в системе искусственного интеллекта. Удачным выбором являются базы данных с сайта: <http://allexcel.ru/gotovye-tablitsy-excel-besplatno>. В данной работе использована база данных: «Таблица базы данных сотрудников, расчет выплат». При этом наиболее достоверной в данном приложении оказались модели INF4, основанная на семантической мере целесообразности информации А.Харкевича при интегральном критерии «Сумма знаний». Точность модели составляет 0,960, что заметно выше, чем достоверность экспертных оценок, которая считается равной около 70%. Для оценки достоверности моделей в АСК-анализе и системе «Эйдос» используется F-критерий Ван Ризбергена и его нечеткое мультиклассовое обобщение, предложенное проф.Е.В.Луценко

The creation of artificial intelligence systems is one of important and perspective directions of development of modern information technology. As there are many alternatives to artificial intelligence systems, there is a need to evaluate mathematical models of these systems. In this work, we consider a solution of the problem of identifying classes of levels of pay of employees on their characteristics. To achieve this goal, it requires free access to test the source data and methodology, which will help to convert the data into the form needed for work in artificial intelligence systems. A good choice is the databases from the site: <http://allexcel.ru/gotovye-tablitsy-excel-besplatno>. In this work, we have used the database called "The database table of employees, payments calculation". The most reliable in this application was the model of the INF4 based on semantic appropriate measure of information of A. Kharkevich with integral criteria of "Amount of knowledge". The accuracy of the model is 0.960, which is much higher than the reliability of expert evaluations, which is equal to about 70%. To assess the reliability of the models in the ACS-analysis and the system called "Eidos" we have used F-criterion of van Ritbergen and fuzzy multiclass generalization proposed by Professor E. V. Lutsenko

Ключевые слова: ИССЛЕДОВАНИЕ ЗАВИСИМОСТИ ОПЛАТЫ СОТРУДНИКОВ АПК ОТ ИХ ХАРАКТЕРИСТИК

Keywords: INVESTIGATION OF THE DEPENDENCE OF PAYMENT TO EMPLOYEES OF AIC FROM THEIR CHARACTERISTICS

Doi: 10.21515/1990-4665-124-002

СОДЕРЖАНИЕ

ВВЕДЕНИЕ	2
1. СИНТЕЗ И ВЕРИФИКАЦИЯ МОДЕЛЕЙ	3
1.1. ОПИСАНИЕ РЕШЕНИЯ.....	3
1.2. ПРЕОБРАЗОВАНИЕ ИСХОДНЫХ ДАННЫХ ИЗ HTML-ФОРМАТА В ФАЙЛ ИСХОДНЫХ ДАННЫХ MS EXCEL 3	
1.3. СИНТЕЗ И ВЕРИФИКАЦИЯ СТАТИСТИЧЕСКИХ И ИНТЕЛЛЕКТУАЛЬНЫХ МОДЕЛЕЙ	1
1.4. ВИДЫ МОДЕЛЕЙ СИСТЕМЫ «ЭЙДОС»	3
1.5. РЕЗУЛЬТАТЫ ВЕРИФИКАЦИИ МОДЕЛЕЙ	5
2. РЕШЕНИЕ ЗАДАЧ ИДЕНТИФИКАЦИИ В НАИБОЛЕЕ ДОСТОВЕРНОЙ МОДЕЛИ	11
2.1. РЕШЕНИЕ ЗАДАЧИ.....	11
2.2. КОГНИТИВНЫЕ ФУНКЦИИ	14
2.3. SWOT и PEST МАТРИЦЫ И ДИАГРАММЫ	18
2.4. КЛАСТЕРНО-КОНСТРУКТИВНЫЙ АНАЛИЗ ПРИЗНАКОВ	21
ЗАКЛЮЧЕНИЕ.....	22
ЛИТЕРАТУРА.....	23

ВВЕДЕНИЕ

Создание систем искусственного интеллекта является одним из важных и перспективных направлений развития современных информационных технологий. Так как существует множество альтернатив систем искусственного интеллекта, то возникает необходимость оценки качества математических моделей этих систем. В данной работе рассмотрено решение задачи идентификации классов уровней оплаты сотрудников фирмы по их характеристикам.

Для достижения поставленной цели необходимы свободный доступ к тестовым исходным данным и методика, которая поможет преобразовать эти данные в форму, которая необходима для работы в системе искусственного интеллекта. Удачным выбором является база данных тестовых задач для систем искусственного интеллекта репозитория UCI.

В данной работе использована база данных «baza dannix sotrudnikov» из банка исходных данных по задачам искусственного интеллекта – репозитория UCI.

Для решения задачи используем стандартные возможности Microsoft Office Word и Excel, блокнот, а также систему искусственного интеллекта "Эйдос- X++".

1. СИНТЕЗ И ВЕРИФИКАЦИЯ МОДЕЛЕЙ

1.1. Описание решения

В соответствии с методологией АСК-анализа решение поставленной задачи проведем в четыре этапа:

1. Преобразование исходных данных из HTML-формата в промежуточные файлы MS Excel.
2. Преобразование исходных данных из промежуточных файлов MS Excel в базы данных системы "Эйдос".
3. Синтез и верификация моделей предметной области.
4. Применение моделей для решения задач идентификации, прогнозирования и исследования предметной области.

1.2. Преобразование исходных данных из HTML-формата в файл исходных данных MS Excel

Из электронного ресурса баз данных <http://allexcel.ru/gotovye-tablitsy-excel-besplatno> возьмем базу данных сотрудников – «Таблица базы данных сотрудников, расчет выплат»: <http://allexcel.ru/images/Excel/baza%20dannix%20sotrudnikov2.xls>, в которой оставим следующие колонки:

1. Должность
2. Образование
3. Пол
4. Год рождения
5. Год приема на работу
6. Количество детей

7. Стаж
8. Номер отдела
9. Доход

Столбцы 1-8 описательные шкалы.

Столбец 9 является классификационной шкалой. Этот столбец показывает месячный доход сотрудников.

Обучающая выборка:

Таблица 1 – baza dannix sotrudnikov.xls

п	Должность	Образование	Пол	Год рожд.	Год приема на работу	Кол- во детей	Стаж	№ отдела	Доход
1	секретарь	среднее	ж	1974	2001	3	13	1	51660
2	секретарь	среднее	ж	1988	2007	0	7	1	50400
3	секретарь референт	высшее	ж	1979	1995	2	19	1	57270
4	юрист	высшее	м	1947	1991	1	23	1	93375
5	юрист	высшее	м	1978	1997	0	17	1	93375
6	юрист	высшее	м	1988	2011	2	3	1	90000
7	бухгалтер	высшее	ж	1970	1987	1	27	2	49800
8	бухгалтер	среднее	ж	1951	1995	4	19	2	49800
9	бухгалтер	среднее спец.	ж	1953	2001	1	13	2	49200
10	экономист	высшее	ж	1962	1990	3	24	2	54780
11	экономист	высшее	ж	1984	1996	1	18	2	54780
12	главный бухгалтер	высшее	м	1964	2000	2	14	2	58080
13	бухгалтер	высшее	ж	1954	1994	1	20	3	64740
14	бухгалтер	высшее	ж	1978	2011	0	3	3	62400
15	бухгалтер	высшее	м	1949	2002	3	12	3	62920
16	дендролог	высшее	ж	1966	2007	1	7	3	66000
17	архитектор	высшее	м	1981	1997	1	17	3	99600
18	экономист	высшее	м	1978	2006	3	8	3	98400
19	экономист	высшее	ж	1952	2001	2	13	3	100860
20	рабочий	среднее	ж	1988	2005	0	9	4	45600
21	рабочий	среднее	м	1982	1995	1	19	4	47310
22	рабочий	среднее	м	1964	1993	3	21	4	47310
23	рабочий	среднее	м	1981	1997	1	17	4	47310
24	рабочий	среднее	м	1967	1991	3	23	4	47310
25	рабочий	среднее	м	1983	2003	1	11	4	45980
26	рабочий	среднее	ж	1983	2000	0	14	4	46740
27	рабочий	среднее	м	1998	2012	0	2	4	45600
28	водитель	среднее	м	1949	1991	2	23	4	48555
29	водитель	среднее спец.	м	1967	2006	4	8	4	46800
30	водитель	среднее	м	1981	2001	1	13	4	47190
31	агроном	среднее спец.	ж	1971	2007	0	7	4	50400
32	агроном	среднее спец.	ж	1954	1990	3	24	4	52290
33	инженер	высшее	м	1983	2013	0	1	4	59250
34	инженер	высшее	м	1992	2012	3	2	4	52800
35	инженер	высшее	м	1965	1992	1	22	4	61230
36	инженер	высшее	м	1981	1992	1	22	4	61230

37	лаборант	среднее	ж	1990	2007	0	7	5	43200
38	лаборант	среднее	ж	1991	2007	0	7	5	43200
39	лаборант	среднее	ж	1989	2001	3	13	5	44280
40	лаборант	среднее	ж	1982	2001	1	13	5	44280
41	техник	среднее	м	1989	2006	1	8	5	48000
42	техник	среднее спец.	м	1979	1995	2	19	5	56250
43	техник	среднее спец.	м	1993	2009	0	5	5	48000
44	инженер	высшее	м	1969	1997	3	17	5	58740
45	инженер	высшее	м	1968	1992	1	22	5	58740

Поскольку ввод исходных данных в систему «Эйдос» планируется осуществить с помощью ее универсального программного интерфейса импорта данных из внешних баз данных, который работает с файлами MS Excel, то преобразуем данные из html-файла в xls-файл, для чего выполним следующие операции.

Скопируем получившуюся таблицу из MS Word в MS Excel и запишем ее с именем: Inp_data.xls в папку: c:\Aidos-X\AID_DATA\Inp_data\. В файле Inp_data.xls добавим пустую колонку на позиции «А» и автоматически пронумеруем все строки. В результате получим таблицу исходных данных, полностью подготовленную для обработки в системе «Эйдос» и записанную в нужную папку в виде файла нужного типа с нужным именем.

Автоматизированная формализация предметной области путем импорта исходных данных из внешних баз данных в систему "Эйдос".

Для загрузки базы исходных данных в систему «Эйдос» необходимо воспользоваться универсальным программным интерфейсом для ввода данных из внешних баз данных табличного вида, т.е. режимом 2.3.2.2 (рисунок 1).

Рисунок 1. Экранная форма Универсального программного интерфейса импорта данных в систему "Эйдос" (режим 2.3.2.2.)

В экранной форме, приведенной на рисунке 5, задать настройки, показанные на рисунке:

- "Задайте тип файла исходных данных Inp_data": "XLS – MS Excel-2003";
- "Задайте диапазон столбцов классификационных шкал": "Начальный столбец классификационных шкал" – 10, "Конечный столбец классификационных шкал" – 10(последний столбец в таблице);

– "Задайте диапазон столбцов описательных шкал": "Начальный столбец описательных шкал" – 2, "Конечный столбец описательных шкал" – 9;

– "Задание параметров формирования сценариев или способа интерпретации текстовых полей": "Не применять сценарный метод АСК-анализа и спец.интерпретацию ТХТ-полей".

После нажать кнопку "ОК". Далее открывается окно, где размещена информация о размерности модели (рисунок 2). В этом окне необходимо нажать кнопку "Выйти на создание модели".

2.3.2.2. Задание размерности модели системы "Эйдос-Х++"

ЗАДАНИЕ В ДИАЛОГЕ РАЗМЕРНОСТИ МОДЕЛИ

Суммарное количество градаций классификационных и описательных шкал: [4 x 39]

Тип шкалы	Количество классификационных шкал	Количество градаций классификационных	Среднее количество градаций на класс. шкалу	Количество описательных шкал	Количество градаций описательных шкал	Среднее количество градаций на опис. шкалу
Числовые	1	4	4,00	5	20	4,00
Текстовые	0	0	0,00	3	19	6,33
ВСЕГО:	1	4	4,00	8	39	4,87

Задайте число интервалов (градаций) в шкале:

В классификационных шкалах: В описательных шкалах:

Пересчитать шкалы и градации Параметры числ. шкал и градаций Выйти на создание модели

Рисунок 2. Задание размерности модели системы "Эйдос"

Далее открывается окно, отображающее стадию процесса импорта данных из внешней БД "Inp_data.xls" в систему "Эйдос" (рисунок 3), а также прогноз времени завершения этого процесса. В том окне необходимо дождаться завершения формализации предметной области и нажать кнопку "ОК".

Рисунок 3. Процесс импорта данных из внешней БД "Inp_data.xls" в систему "Эйдос"

В результате формируются классификационные и описательные шкалы и градации, с применением которых исходные данные кодируются и представляются в форме эвентологических баз данных. Этим самым полностью автоматизировано выполняется 2-й этап АСК-анализа «Формализация предметной области». Для просмотра классификационных шкал и градаций необходимо запустить режим 2.1 (рисунок 4).

Рисунок 6. Классификационные шкалы и градации

Для просмотра описательных шкал и градаций необходимо запустить режим 2.2 (рисунок 5), а обучающей выборки режим 2.3.1. (рисунок 6):

Рисунок 5. Описательные шкалы и градации (фрагмент)

Рисунок 6. Обучающая выборка (фрагмент)

ПАРАМЕТРЫ ШКАЛ И ГРАДАЦИЙ С АДАПТИВНЫМИ ГРАНИЦАМИ И ПРИМЕРНО РАВНЫМ КОЛИЧЕСТВОМ НАБЛЮДЕНИЙ ПО ГРАДАЦИЯМ с коррекцией ошибки округления числа наблюдений по интервалу градации при переходе к следующей градации

КЛАССИФИКАЦИОННАЯ ШКАЛА: код: [1], наим.: "ДОХОД", набл.на шкалу (всего): 45, тип/число градаций в шкале: "Равное число событий в интервалах"/4
 1 Наим.градации: 1/4-{ 43200.0000000, 47310.0000000}, размер интервала= 4110.0000000, расч./факт.число наблюдений на градацию: 11/11
 2 Наим.градации: 2/4-{ 47310.0000000, 50400.0000000}, размер интервала= 3090.0000000, расч./факт.число наблюдений на градацию: 11/11
 3 Наим.градации: 3/4-{ 50400.0000000, 59250.0000000}, размер интервала= 8850.0000000, расч./факт.число наблюдений на градацию: 11/11
 4 Наим.градации: 4/4-{ 59250.0000000, 100860.0000000}, размер интервала= 41610.0000000, расч./факт.число наблюдений на градацию: 12/12

ОПИСАТЕЛЬНАЯ ШКАЛА: код: [1], наим.: "ДОЛЖНОСТЬ", тип/число градаций в шкале: "Равное число событий в интервалах"/14

- 1 Наим.градации: 1/14-агроном
- 2 Наим.градации: 2/14-архитектор
- 3 Наим.градации: 3/14-бухгалтер
- 4 Наим.градации: 4/14-водитель
- 5 Наим.градации: 5/14-главный бухгалтер
- 6 Наим.градации: 6/14-дендролог
- 7 Наим.градации: 7/14-инженер
- 8 Наим.градации: 8/14-лаборант
- 9 Наим.градации: 9/14-рабочий
- 10 Наим.градации: 10/14-секретарь
- 11 Наим.градации: 11/14-секретарь референт
- 12 Наим.градации: 12/14-техник
- 13 Наим.градации: 13/14-экономист
- 14 Наим.градации: 14/14-юрист

ОПИСАТЕЛЬНАЯ ШКАЛА: код: [2], наим.: "ОБРАЗОВАНИЕ", тип/число градаций в шкале: "Равное число событий в интервалах"/3

- 15 Наим.градации: 1/3-высшее
- 16 Наим.градации: 2/3-среднее
- 17 Наим.градации: 3/3-среднее спец.

ОПИСАТЕЛЬНАЯ ШКАЛА: код: [3], наим.: "ПОЛ", тип/число градаций в шкале: "Равное число событий в интервалах"/2

- 18 Наим.градации: 1/2-ж
- 19 Наим.градации: 2/2-м

ОПИСАТЕЛЬНАЯ ШКАЛА: код: [4], наим.: "ГОД РОЖД.", набл.на шкалу (всего): 45, тип/число градаций в шкале: "Равное число событий в интервалах"/4

- 20 Наим.градации: 1/4-{1947.0000000, 1964.0000000}, размер интервала= 17.0000000, расч./факт.число наблюдений на градацию: 11/11
- 21 Наим.градации: 2/4-{1964.0000000, 1978.0000000}, размер интервала= 14.0000000, расч./факт.число наблюдений на градацию: 11/11
- 22 Наим.градации: 3/4-{1978.0000000, 1983.0000000}, размер интервала= 5.0000000, расч./факт.число наблюдений на градацию: 11/11
- 23 Наим.градации: 4/4-{1983.0000000, 1998.0000000}, размер интервала= 15.0000000, расч./факт.число наблюдений на градацию: 12/12

ОПИСАТЕЛЬНАЯ ШКАЛА: код: [5], наим.: "ГОД ПРИЕМА НА РАБОТУ", набл.на шкалу (всего): 45, тип/число градаций в шкале: "Равное число событий в интервалах"/4

- 24 Наим.градации: 1/4-{1987.0000000, 1994.0000000}, размер интервала= 7.0000000, расч./факт.число наблюдений на градацию: 11/11
- 25 Наим.градации: 2/4-{1994.0000000, 2000.0000000}, размер интервала= 6.0000000, расч./факт.число наблюдений на градацию: 11/11
- 26 Наим.градации: 3/4-{2000.0000000, 2006.0000000}, размер интервала= 6.0000000, расч./факт.число наблюдений на градацию: 11/11
- 27 Наим.градации: 4/4-{2006.0000000, 2013.0000000}, размер интервала= 7.0000000, расч./факт.число наблюдений на градацию: 12/12

ОПИСАТЕЛЬНАЯ ШКАЛА: код: [6], наим.: "КОЛ-ВО ДЕТЕЙ", набл.на шкалу (всего): 34, тип/число градаций в шкале: "Равное число событий в интервалах"/4

- 28 Наим.градации: 1/4-{1.0000000, 1.0000000}, размер интервала=0.0000000, расч./факт.число наблюдений на градацию: 8/8
- 29 Наим.градации: 2/4-{1.0000000, 1.0000000}, размер интервала=0.0000000, расч./факт.число наблюдений на градацию: 8/8
- 30 Наим.градации: 3/4-{1.0000000, 3.0000000}, размер интервала=2.0000000, расч./факт.число наблюдений на градацию: 9/9
- 31 Наим.градации: 4/4-{3.0000000, 4.0000000}, размер интервала=1.0000000, расч./факт.число наблюдений на градацию: 9/9

ОПИСАТЕЛЬНАЯ ШКАЛА: код: [7], наим.: "СТАЖ", набл.на шкалу (всего): 45, тип/число градаций в шкале: "Равное число событий в интервалах"/4

- 32 Наим.градации: 1/4-{ 1.0000000, 7.0000000}, размер интервала= 6.0000000, расч./факт.число наблюдений на градацию: 11/11
- 33 Наим.градации: 2/4-{ 7.0000000, 13.0000000}, размер интервала= 6.0000000, расч./факт.число наблюдений на градацию: 11/11
- 34 Наим.градации: 3/4-{13.0000000, 19.0000000}, размер интервала= 6.0000000, расч./факт.число наблюдений на градацию: 11/11
- 35 Наим.градации: 4/4-{19.0000000, 27.0000000}, размер интервала= 8.0000000, расч./факт.число наблюдений на градацию: 12/12

ОПИСАТЕЛЬНАЯ ШКАЛА: код: [8], наим.: "№ ОТДЕЛА", набл.на шкалу (всего): 45, тип/число градаций в шкале: "Равное число событий в интервалах"/4

- 36 Наим.градации: 1/4-{1.0000000, 2.0000000}, размер интервала=1.0000000, расч./факт.число наблюдений на градацию: 11/11
- 37 Наим.градации: 2/4-{2.0000000, 4.0000000}, размер интервала=2.0000000, расч./факт.число наблюдений на градацию: 11/11
- 38 Наим.градации: 3/4-{4.0000000, 4.0000000}, размер интервала=0.0000000, расч./факт.число наблюдений на градацию: 11/11
- 39 Наим.градации: 4/4-{4.0000000, 5.0000000}, размер интервала=1.0000000, расч./факт.число наблюдений на градацию: 12/12

Тем самым создаются все необходимые и достаточные предпосылки для выявления силы и направления причинно-следственных связей между значениями факторов и результатами их совместного системного воздействия (с учетом нелинейности системы [11]).

1.3. Синтез и верификация статистических и интеллектуальных моделей

Далее запускаем режим 3.5, в котором задаются модели для синтеза и верификации, а также задается модель, которой по окончании режима присваивается статус текущей (рисунок 7).

Рисунок 7. Выбор моделей для синтеза и верификации, а также текущей модели

В данном режиме имеется много различных методов верификации моделей, в том числе и поддерживающие бутстрепный метод. Но мы используем параметры по умолчанию, приведенные на рисунке 7. Стадия процесса исполнения режима 3.5 и прогноз времени его окончания отображаются на экранной форме, приведенной на рисунке 8.

Рисунок 8. Синтез и верификация статистических моделей и моделей знаний

Интересно заметить (см. рисунок 8), что синтез и верификация всех 10 моделей на данной задаче заняли 18 секунд. При этом верификация (оценка достоверности моделей) проводилась на всех 45 примерах наблюдения из обучающей выборки. В результате выполнения режима 3.5 созданы все модели, со всеми частными критериями, перечисленные на рисунке 10, но ниже мы приведем лишь некоторые из них (таблицы 2, 3, 4).

1.4. Виды моделей системы «Эйдос»

Рассмотрим решение задачи идентификации на примере модели INF1, в которой рассчитано количество информации по А.Харкевичу, которое мы получаем о принадлежности идентифицируемого объекта к каждому из классов, если знаем, что у этого объекта есть некоторый признак.

По сути, частные критерии представляют собой просто формулы для преобразования матрицы абсолютных частот (таблица 2) в матрицы условных и безусловных процентных распределений, и матрицы знаний (таблицы 3 и 4).

Таблица 2 – Матрица абсолютных частот (модель ABS) и условных и безусловных процентных распределений (фрагменты)

5.5. Модель: "1. ABS - частный критерий: количество встреч сочетаний: "Класс-признак" у объектов обуч.выборки"

Код признака	Наименование описательной шкалы и градации	1. ДОХОД 1/4 (43200.00... 47310.000...	2. ДОХОД 2/4 (47310.00... 50400.000...	3. ДОХОД 3/4 (50400.00... 59250.000...	4. ДОХОД 4/4 (59250.00... 100860.00...	Сумма	Среднее	Средн. квадр. откл.
1	ДОЛЖНОСТЬ-1/14-агроном		1	1		2	0.500	0.577
2	ДОЛЖНОСТЬ-2/14-архитектор				1	1	0.250	0.500
3	ДОЛЖНОСТЬ-3/14-бухгалтер		3		3	6	1.500	1.732
4	ДОЛЖНОСТЬ-4/14-водитель	2	1			3	0.750	0.957
5	ДОЛЖНОСТЬ-5/14-главный бухгалтер			1		1	0.250	0.500
6	ДОЛЖНОСТЬ-6/14-дендролог				1	1	0.250	0.500
7	ДОЛЖНОСТЬ-7/14-инженер			4	2	6	1.500	1.915
8	ДОЛЖНОСТЬ-8/14-лаборант	4				4	1.000	2.000
9	ДОЛЖНОСТЬ-9/14-рабочий	8				8	2.000	4.000
10	ДОЛЖНОСТЬ-10/14-секретарь		1	1		2	0.500	0.577
11	ДОЛЖНОСТЬ-11/14-секретарь референт			1		1	0.250	0.500
12	ДОЛЖНОСТЬ-12/14-техник		2	1		3	0.750	0.957
13	ДОЛЖНОСТЬ-13/14-экономист			2	2	4	1.000	1.155
14	ДОЛЖНОСТЬ-14/14-юрист				3	3	0.750	1.500
15	ОБРАЗОВАНИЕ-1/3-высшее		1	8	12	21	5.250	5.737
16	ОБРАЗОВАНИЕ-2/3-среднее	13	4	1		18	4.500	5.916
17	ОБРАЗОВАНИЕ-3/3-среднее спец.	1	3	2		6	1.500	1.291
18	ПОЛ-1/2-ж	6	5	5	4	20	5.000	0.816

Помощь MS Excel MS Word

Таблица 3 – Матрица информативностей (модель INF1) в битах (фрагмент)

5.5. Модель: "4. INF1 - частный критерий: количество знаний по А.Харкевичу; вероятности из PRC1"

Код признака	Наименование описательной шкалы и градации	1. ДОХОД 1/4 (43200.00... 47310.000...	2. ДОХОД 2/4 (47310.00... 50400.000...	3. ДОХОД 3/4 (50400.00... 59250.000...	4. ДОХОД 4/4 (59250.00... 100860.00...	Сумма	Среднее	Средн. кв.откл.
1	ДОЛЖНОСТЬ-1/14-агроном		0.359	0.238		0.597	0.149	0.179
2	ДОЛЖНОСТЬ-2/14-архитектор				0.448	0.448	0.112	0.224
3	ДОЛЖНОСТЬ-3/14-бухгалтер		0.359		0.211	0.570	0.143	0.175
4	ДОЛЖНОСТЬ-4/14-водитель	0.265	0.221			0.486	0.121	0.141
5	ДОЛЖНОСТЬ-5/14-главный бухгалтер	...		0.475		0.475	0.119	0.237
6	ДОЛЖНОСТЬ-6/14-дендролог				0.448	0.448	0.112	0.224
7	ДОЛЖНОСТЬ-7/14-инженер			0.336	0.073	0.409	0.102	0.160
8	ДОЛЖНОСТЬ-8/14-лаборант	0.404				0.404	0.101	0.202
9	ДОЛЖНОСТЬ-9/14-рабочий	0.404				0.404	0.101	0.202
10	ДОЛЖНОСТЬ-10/14-секретарь		0.359	0.238		0.597	0.149	0.179
11	ДОЛЖНОСТЬ-11/14-секретарь референт	...		0.475		0.475	0.119	0.237
12	ДОЛЖНОСТЬ-12/14-техник		0.457	0.099		0.557	0.139	0.217
13	ДОЛЖНОСТЬ-13/14-экономист			0.238	0.211	0.449	0.112	0.130
14	ДОЛЖНОСТЬ-14/14-юрист				0.448	0.448	0.112	0.224
15	ОБРАЗОВАНИЕ-1/3-высшее		-0.444	0.145	0.257	-0.042	-0.011	0.308
16	ОБРАЗОВАНИЕ-2/3-среднее	0.293	0.082	-0.513		-0.138	-0.035	0.342
17	ОБРАЗОВАНИЕ-3/3-среднее спец.	-0.208	0.359	0.099		0.250	0.063	0.236
18	ПОЛ-1/2-ж	-0.007	0.122	0.001	-0.102	0.014	0.004	0.092

Помощь MS Excel MS Word

Таблица 4 – Матрица знаний (модель INF3) (фрагмент)

5.5. Модель: "6. INF3 - частный критерий: Хи-квадрат, разности между фактическими и ожидаемыми абс.частотами"

Код признака	Наименование описательной шкалы и градации	1. ДОХОД 1/4 (43200.00... 47310.000...	2. ДОХОД 2/4 (47310.00... 50400.000...	3. ДОХОД 3/4 (50400.00... 59250.000...	4. ДОХОД 4/4 (59250.00... 100860.00...	Сумма	Среднее	Средн. кв.откл.
1	ДОЛЖНОСТЬ-1/14-агроном	-0.613	0.650	0.501	-0.539			0.668
2	ДОЛЖНОСТЬ-2/14-архитектор	-0.307	-0.175	-0.249	0.731			0.490
3	ДОЛЖНОСТЬ-3/14-бухгалтер	-1.840	1.951	-1.496	1.384			1.945
4	ДОЛЖНОСТЬ-4/14-водитель	1.080	0.476	-0.748	-0.808			0.932
5	ДОЛЖНОСТЬ-5/14-главный бухгалтер	...	-0.307	0.751	-0.269			0.504
6	ДОЛЖНОСТЬ-6/14-дендролог	-0.307	-0.175	-0.249	0.731			0.490
7	ДОЛЖНОСТЬ-7/14-инженер	-1.840	-1.049	2.504	0.384			1.906
8	ДОЛЖНОСТЬ-8/14-лаборант	2.774	-0.699	-0.997	-1.077			1.856
9	ДОЛЖНОСТЬ-9/14-рабочий	5.547	-1.398	-1.994	-2.155			3.712
10	ДОЛЖНОСТЬ-10/14-секретарь	-0.613	0.650	0.501	-0.539			0.668
11	ДОЛЖНОСТЬ-11/14-секретарь референт	...	-0.307	0.751	-0.269			0.504
12	ДОЛЖНОСТЬ-12/14-техник	-0.920	1.476	0.252	-0.808			1.117
13	ДОЛЖНОСТЬ-13/14-экономист	-1.226	-0.699	1.003	0.923			1.133
14	ДОЛЖНОСТЬ-14/14-юрист	-0.920	-0.524	-0.748	2.192			1.470
15	ОБРАЗОВАНИЕ-1/3-высшее	-6.438	-2.670	2.765	6.344			5.671
16	ОБРАЗОВАНИЕ-2/3-среднее	7.481	0.854	-3.487	-4.848			5.549
17	ОБРАЗОВАНИЕ-3/3-среднее спец.	-0.840	1.951	0.504	-1.616			1.568
18	ПОЛ-1/2-ж	-0.132	1.504	0.014	-1.387			1.184

Помощь MS Excel MS Word

1.5. Результаты верификации моделей

Результаты верификации (оценки достоверности) моделей, отличающихся частными критериями с двумя приведенными выше интегральными критериями приведены на рисунке 9.

4.13.6. Обобщенная форма для достоверности моделей при разнотипном критерии. Текущая модель: "INF1"

Наименование модели и частного критерия	Интегральный критерий	Всего логических объектов выборки	Число истинно-положительных решений (TP)	Число истинно-отрицательных решений (TN)	Число ложно-положительных решений (FP)	Число ложно-отрицательных решений (FN)	Точность модели	Полнота модели	Ф-мера Ван Ризбергера	Сумма модулей уровней схожести истинно-положительных решений (ST)	Сумма модулей уровней схожести истинно-отрицательных решений (ST)	Сумма модулей уровней схожести ложно-положительных решений (SFP)	Сумма модулей уровней схожести ложно-отрицательных решений (SF)	S-T
1. ABS - частный критерий: количество встреч сочетаний "класс..."	Корреляция абсолютных частот с...	45	45	24	111		0,288	1,000	0,448	29,502	1,922	30,956		
2. PRC1 - частный критерий: условная вероятность его признака сред...	Корреляция условных частот с о...	45	45	24	111		0,288	1,000	0,448	29,502	1,922	30,956		
3. PRC2 - частный критерий: условная вероятность его признака сред...	Корреляция условных частот с о...	45	45	24	111		0,288	1,000	0,448	29,502	1,922	30,956		
4. INF1 - частный критерий: количество знаний по А/Харкевичу; в...	Семантический резонанс зна...	45	37	118	17	8	0,685	0,822	0,747	14,081	45,112	2,376	1,300	
5. INF2 - частный критерий: количество знаний по А/Харкевичу; в...	Семантический резонанс зна...	45	42	93	42	3	0,500	0,933	0,651	20,408	27,802	7,489	0,574	
6. INF3 - частный критерий: количество знаний по А/Харкевичу; в...	Семантический резонанс зна...	45	37	119	16	8	0,685	0,822	0,755	14,124	45,381	2,379	1,315	
7. INF4 - частный критерий: Уникальность, разности между факти...	Семантический резонанс зна...	45	43	97	38	2	0,531	0,956	0,683	26,308	37,846	8,486	0,305	
8. INF5 - частный критерий: ROI (Return On Investment); веротно...	Семантический резонанс зна...	45	43	97	38	2	0,531	0,956	0,683	26,316	6,023	0,159		
9. INF6 - частный критерий: ROI (Return On Investment); веротно...	Семантический резонанс зна...	45	37	123	12	8	0,755	0,822	0,787	16,705	55,380	1,345	1,216	
10. INF7 - частный критерий: разнотипный и безупречный; вер...	Семантический резонанс зна...	45	45	66	69		0,395	1,000	0,566	25,005	9,570	11,861		
11. INF8 - частный критерий: ROI (Return On Investment); веротно...	Семантический резонанс зна...	45	37	123	12	8	0,755	0,822	0,787	16,754	55,798	1,333	1,199	
12. INF9 - частный критерий: ROI (Return On Investment); веротно...	Семантический резонанс зна...	45	45	60	75		0,375	1,000	0,545	23,954	9,201	11,355		
13. INF10 - частный критерий: разнотипный и безупречный; вер...	Семантический резонанс зна...	45	42	100	35	3	0,545	0,933	0,689	23,191	33,534	7,981	0,579	
14. INF11 - частный критерий: разнотипный и безупречный; вер...	Семантический резонанс зна...	45	43	73	62	2	0,410	0,956	0,573	23,306	16,119	12,298	0,259	
15. INF12 - частный критерий: разнотипный и безупречный; вер...	Семантический резонанс зна...	45	42	101	34	3	0,553	0,933	0,694	23,482	33,981	8,161	0,570	
16. INF13 - частный критерий: разнотипный и безупречный; вер...	Семантический резонанс зна...	45	43	77	58	2	0,426	0,956	0,589	24,239	16,894	12,907	0,266	

а)

4.13.6. Обобщенная форма для достоверности моделей при разнотипном критерии. Текущая модель: "INF1"

Число истинно-отрицательных решений (TN)	Число ложно-положительных решений (FP)	Число ложно-отрицательных решений (FN)	Точность модели	Полнота модели	Ф-мера Ван Ризбергера	Сумма модулей уровней схожести истинно-положительных решений (ST)	Сумма модулей уровней схожести истинно-отрицательных решений (ST)	Сумма модулей уровней схожести ложно-положительных решений (SFP)	Сумма модулей уровней схожести ложно-отрицательных решений (SF)	S-Точность модели	S-Полнота модели	Л-мера пров. Е.В. / учен.	Процент правильной идентификации	Процент правильной не идентификации	Процент ошибочной идентификации	Процент ошибочной не идентификации	Процент правильного результата	Дата получения результата	Время получения результата
24	111		0,288	1,000	0,448	29,502	1,922	30,956	0,488	1,000	0,656	100,000	100,000	19,199	80,801		59,599	09.12.2016	23:00:22
135			0,250	1,000	0,400	28,177		50,387	0,359	1,000	0,528	100,000	100,000				50,000	09.12.2016	23:00:22
24	111		0,288	1,000	0,448	29,502	1,922	30,956	0,488	1,000	0,656	100,000	100,000	19,199	80,801		59,599	09.12.2016	23:00:23
135			0,250	1,000	0,400	32,774		62,602	0,344	1,000	0,511	100,000	100,000				50,000	09.12.2016	23:00:23
24	111		0,288	1,000	0,448	29,502	1,922	30,956	0,488	1,000	0,656	100,000	100,000	19,199	80,801		59,599	09.12.2016	23:00:24
135			0,250	1,000	0,400	33,258		63,521	0,344	1,000	0,512	100,000	100,000				50,000	09.12.2016	23:00:24
118	17	8	0,685	0,822	0,747	14,081	45,112	2,376	1,300	0,856	0,915	0,885	82,222	86,685	13,315	17,778	84,453	09.12.2016	23:00:25
93	42	3	0,500	0,933	0,651	20,408	27,802	7,489	0,574	0,732	0,973	0,835	93,333	68,486	31,514	6,667	80,909	09.12.2016	23:00:25
119	16	8	0,698	0,822	0,755	14,124	45,381	2,379	1,315	0,856	0,915	0,884	82,222	87,165	12,835	17,778	84,694	09.12.2016	23:00:26
93	42	3	0,500	0,933	0,651	21,010	28,726	7,784	0,609	0,730	0,972	0,834	93,333	68,770	31,230	6,667	81,052	09.12.2016	23:00:26
97	38	2	0,531	0,956	0,683	26,308	37,846	8,486	0,305	0,756	0,989	0,857	95,556	71,827	28,173	4,444	83,692	09.12.2016	23:00:27
97	38	2	0,531	0,956	0,683	20,452	26,316	6,023	0,159	0,772	0,992	0,869	95,556	71,827	28,173	4,444	83,692	09.12.2016	23:00:27
123	12	8	0,755	0,822	0,787	16,705	55,380	1,345	1,216	0,925	0,932	0,929	82,222	91,444	8,556	17,778	86,833	09.12.2016	23:00:28
66	69		0,395	1,000	0,566	25,005	9,570	11,861	0,678	1,000	0,808	100,000	48,391	51,609			74,196	09.12.2016	23:00:28
123	12	8	0,755	0,822	0,787	16,754	55,798	1,333	1,199	0,926	0,933	0,930	82,222	91,444	8,556	17,778	86,833	09.12.2016	23:00:29
60	75		0,375	1,000	0,545	23,954	9,201	11,355	0,678	1,000	0,808	100,000	43,661	56,339			71,830	09.12.2016	23:00:30
100	35	3	0,545	0,933	0,689	23,191	33,534	7,981	0,579	0,744	0,976	0,844	93,333	72,326	27,674	6,667	82,830	09.12.2016	23:00:31
73	62	2	0,410	0,956	0,573	23,306	16,119	12,298	0,259	0,655	0,989	0,788	95,556	51,202	48,798	4,444	73,379	09.12.2016	23:00:31
101	34	3	0,553	0,933	0,694	23,482	33,981	8,161	0,570	0,742	0,976	0,843	93,333	73,329	26,671	6,667	83,331	09.12.2016	23:00:32
77	58	2	0,426	0,956	0,589	24,239	16,894	12,907	0,266	0,653	0,989	0,786	95,556	53,932	46,068	4,444	74,744	09.12.2016	23:00:32

б)

Рисунок 9. Оценки достоверности моделей

Наиболее достоверной в данном приложении оказались модели INF4, INF5 при интегральном критерии «Семантический резонанс знаний». При этом точность модели составляет 0,755 а полнота модели 0,822, что является неплохими показателями. Таким образом, уровень достоверности прогнозирования с применением модели выше, чем экспертных оценок, достоверность которых считается равной примерно 70%. Для оценки

достоверности моделей в АСК-анализе и системе «Эйдос» используется F-критерий Ван Ризбергена, а также его нечеткое мультиклассовое обобщение, предложенное проф.Е.В.Луценко [16] (рисунок 10).

Рисунок 10. Виды прогнозов и принцип определения достоверности моделей по авторскому варианту метрики, сходной с F-критерием

Также обращает на себя внимание, что статистические модели, как правило, дают значительно более низкую средневзвешенную достоверность идентификации и не идентификации, чем модели знаний, и практически никогда – более высокую. Этим и оправдано применение моделей знаний и интеллектуальных технологий. На рисунке 11 приведены частные распределения уровней сходства и различия для верно и ошибочно идентифицированных и неидентифицированных ситуаций в наиболее достоверной модели INF5.

Из рисунка 11 видно, что:

– наиболее достоверная модель INF5 лучше определяет непринадлежность объекта к классу, чем принадлежность (что видно также из рисунка 9);

– модуль уровня сходства-различия в наиболее достоверной модели INF5 для верно идентифицированных и верно неидентифицированных

объектов значительно выше, чем для ошибочно идентифицированных и ошибочно неидентифицированных. Это верно практически для всего диапазона уровней сходства-различия, кроме небольших по модулю значений в диапазоне от 0 до 15% уровня сходства. Для больших же значений уровней сходства-различия (более 20%) также различие между верно и ошибочно идентифицированными и неидентифицированными ситуациями очень велико.

Рисунок 11. Частное распределение сходства-различия верно и ошибочно идентифицированных и неидентифицированных состояний объекта моделирования в модели INF5

Это означает, что если учитывать не просто сами факты верно и ошибочно идентифицированных и неидентифицированных объектов, но и учитывать уровень сходства-различия, то можно свести на нет ошибочные идентификации и неидентификации и оценить достоверность модели значительно точнее, чем с помощью F-критерия Ван Ризбергена. Эта идея

и положена в основу нечеткого мультиклассового обобщения помощью F-критерия Ван Ризбергена, предложенного проф.Е.В.Луценко (L-мера) [16].

Для наиболее достоверной модели INF5 L-мера равна 0,930 при точности модели 0,926, полноте модели: 0,933 (см. рисунок 9б), что является очень хорошими показателями.

Любые данные о наблюдениях можно считать суммой истинного значения и шума, причем ни первое, ни второе неизвестны. Поэтому имеет смысл сравнить созданные модели с чисто случайными моделями, совпадающими по основным характеристикам. В системе «Эйдос» есть лабораторная работа № 2.01: «Исследование RND-модели при различных объемах выборки». Если данная работа устанавливается при отсутствии текущего приложения, то все параметры создаваемых моделей задаются вручную, если же текущая модель существует, как в нашем случае, то все основные ее параметры определяются автоматически (рисунок 12):

Рисунок 12. Экранная форма управления созданием случайных моделей, совпадающих с текущей по размерностям основных баз данных

На рисунке 13 показано частное распределение сходства-различия верно и ошибочно идентифицированных и неидентифицированных состояний в случайной модели INF5.

Рисунок 13. Частное распределение сходства-различия верно и ошибочно идентифицированных и неидентифицированных состояний в случайной модели INF5

Совершенно очевидное различие частотных распределений уровней сходства-различия верно и ошибочно идентифицированных и неидентифицированных состояний объекта моделирования и случайной модели (рисунки 11 и 13) объясняется тем, что в реальных моделях кроме шума есть также и информация об истинных причинно-следственных взаимосвязях факторов и их значений с одной стороны, и состояниями объекта моделирования, которые ими обуславливаются, с другой стороны. Если же такой информации в модели нет, то и распределение получается типа, приведенного на рисунке 13.

На рисунке 14 приведены данные по достоверности статистических и когнитивных моделей, созданных на основе случайной выборки.

Наименование модели и частного критерия	Интегральный критерий	Всего логических объектов выборки	Число истинно-положительных решений (TP)	Число истинно-отрицательных решений (TN)	Число ложно-положительных решений (FP)	Число ложно-отрицательных решений (FN)	Точность модели	Полнота модели	F-мера Ван Рисбергена	Сенситивность (ST)	Специфичность (ST)	Положительная прогностическая ценность (STP)	Отрицательная прогностическая ценность (STN)	Среднее значение
1. ABS - частный критерий: количество встреч сочетаний "класс"	Корреляция абс частот с абс частот по признаку...	45	44	67	68	1	0.393	0.978	0.561	20.086	11.872	13.981	0.061	
2. PR1 - частный критерий: усл. вероятность i-го признака сред...	Корреляция усл.отн частот с о...	45	44	67	68	1	0.393	0.978	0.561	20.086	11.872	13.981	0.061	
3. PR2 - частный критерий: усл. вероятность i-го признака сред...	Корреляция усл.отн частот по при...	45	45		135		0.250	1.000	0.400	27.341		56.539		
4. INF1 - частный критерий: количество знаний по АХаркевичу: в...	Сенантический резонанс: зна...	45	39	94	41	6	0.488	0.867	0.624	17.839	30.351	7.478	0.853	
5. INF2 - частный критерий: количество знаний по АХаркевичу: в...	Сенантический резонанс: зна...	45	39	98	37	6	0.513	0.867	0.645	13.338	27.210	5.556	0.703	
6. INF3 - частный критерий: количество знаний по АХаркевичу: в...	Сенантический резонанс: зна...	45	39	92	43	6	0.476	0.867	0.614	17.731	30.429	7.487	0.870	
7. INF4 - частный критерий: ROI (Return On Investment): веро...	Сенантический резонанс: зна...	45	39	99	36	6	0.520	0.867	0.650	12.394	25.850	4.970	0.722	
8. INF5 - частный критерий: ROI (Return On Investment): веро...	Сенантический резонанс: зна...	45	42	98	37	3	0.532	0.933	0.677	20.948	27.658	7.251	0.336	
9. INF6 - частный критерий: ROI (Return On Investment): веро...	Сенантический резонанс: зна...	45	42	95	37	3	0.532	0.933	0.677	20.996	27.747	7.106	0.345	
10. INF7 - частный критерий: разн усл и безуслов вероятностей: вер...	Сенантический резонанс: зна...	45	39	102	33	6	0.542	0.867	0.667	18.567	30.265	5.493	1.170	
11. INF8 - частный критерий: разн усл и безуслов вероятностей: вер...	Сенантический резонанс: зна...	45	42	73	62	3	0.404	0.933	0.564	17.298	10.074	9.667	0.208	
12. INF9 - частный критерий: разн усл и безуслов вероятностей: вер...	Сенантический резонанс: зна...	45	39	102	33	6	0.542	0.867	0.667	18.690	30.597	5.504	1.195	
13. INF10 - частный критерий: разн усл и безуслов вероятностей: вер...	Сенантический резонанс: зна...	45	41	74	61	4	0.402	0.911	0.558	19.206	11.512	10.172	0.223	
14. INF11 - частный критерий: разн усл и безуслов вероятностей: вер...	Сенантический резонанс: зна...	45	40	97	38	5	0.513	0.889	0.650	19.022	29.802	7.968	0.897	
15. INF12 - частный критерий: разн усл и безуслов вероятностей: вер...	Сенантический резонанс: зна...	45	42	78	57	3	0.424	0.933	0.583	15.112	11.326	9.338	0.203	
16. INF13 - частный критерий: разн усл и безуслов вероятностей: вер...	Сенантический резонанс: зна...	45	40	97	38	5	0.513	0.889	0.650	20.118	30.754	8.115	0.935	
17. INF14 - частный критерий: разн усл и безуслов вероятностей: вер...	Сенантический резонанс: зна...	45	42	76	59	3	0.416	0.933	0.575	16.965	13.053	9.846	0.204	

Рисунок 14. Достоверность статистических и когнитивных моделей, созданных на основе случайной выборки

На основе сравнения рисунков 14 и 9 можно сделать следующие выводы:

– достоверность лучшей модели INF5 по F-критерию, отражающей реальный объект моделирования, примерно на 41% выше, чем аналогичной случайной модели ($0,787/0,558=1,41$);

– различие между достоверностью статистических моделей и моделей знаний, созданных на основе случайной выборки, значительно меньше, чем у моделей, отражающих реальный объект моделирования;

– в реальных моделях кроме шума есть также и информация об истинных причинно-следственных взаимосвязях факторов и их значений с одной стороны, и состояниями объекта моделирования, которые ими обуславливаются, с другой стороны, причем примерно 1/3 достоверности обусловлена отражением в реальных моделях закономерностей предметной области, а 2/3 достоверности обусловлено наличием шума в исходных данных. На основании этого можно предположить, что в исходных данных уровень сигнала о реальных причинно-следственных связях в моделируемой предметной области примерно в два раза ниже уровня шума.

2. РЕШЕНИЕ ЗАДАЧ ИДЕНТИФИКАЦИИ В НАИБОЛЕЕ ДОСТОВЕРНОЙ МОДЕЛИ

2.1. Решение задачи

В соответствии с технологией АСК-анализа зададим текущей модель INF5 (режим 5.6) (рисунок 15) и проведем пакетное распознавание в режиме 4.2.1. (рисунок 16)

Рисунок 15. Экранные формы режима задания модели в качестве текущей

Рисунок 16. Экранная форма режима пакетного распознавания в текущей модели INF5

В результате пакетного распознавания в текущей модели создается ряд баз данных, которые визуализируются в выходных экранных формах, отражающих результаты решения задачи идентификации и прогнозирования.

Режим 4.1.3 системы «Эйдос» обеспечивает отображение результатов идентификации и прогнозирования в различных формах:

1. Подробно наглядно: "Объект – классы".
2. Подробно наглядно: "Класс – объекты".
3. Итоги наглядно: "Объект – классы".
4. Итоги наглядно: "Класс – объекты".
5. Подробно сжато: "Объект – классы".
6. Обобщенная форма по достоверности моделей при разных интегральных критериях.
7. Обобщенный статистический анализ результатов идентификации по моделям и интегральным критериям.
8. Статистический анализ результатов идентификации по классам, моделям и интегральным критериям.

9. Распознавание уровня сходства при разных моделях и интегральных критериях.

10. Достоверность идентификации классов при разных моделях и интегральных критериях.

Ниже кратко рассмотрим некоторые из них.

На рисунках 17 и 18 приведены примеры прогнозов в наиболее достоверной модели INF5:

2.2. Когнитивные функции

Рассмотрим режим 4.5, в котором реализована возможность визуализации когнитивных функций для любых моделей и любых сочетаний классификационных и описательных шкал (рисунок 19)

Рисунок 19. Экранная форма режима 4.5 системы «Эйдос-X++»
«Визуализация когнитивных функций»

Применительно к задаче, рассматриваемой в данной работе, когнитивная функция показывает, какое количество информации содержится в различных значениях факторов о том, что объект моделирования перейдет в те или иные будущие состояния. Когнитивным функциям посвящено много работ автора 9, но наиболее новой и обобщающей из них является работа. Поэтому здесь не будем

КОГНИТИВНАЯ ФУНКЦИЯ:
 Приложение: Созданное путем ввода данных из БД Inp_data. Это название можно скорректировать
 Модель: InFS - частный критерий: ROI (Return On Investment); вероятности из PRC2

7-я подматрица
 Степень редукции: 0,00
 Частично редукцированных когнитивных функций: 0,00
 Положительной от классов: 0,00
 Положительной от признаков: 0,00
 Негативной от классов: 0,00
 Негативной от признаков: 0,00
 Метод интерполяции: Фильтр тип: 1
 Количество градиент уровней: 32
 Цвет нерасчетных нулей: Отсутствует
 Тип палитры: Цветовая заливка, спектр от син
 Интерполяция нерасчетных нулей: Отсутствует
 Размер шрифта 45

8.№ ОТДЕЛА ПРИЗНАКИ

Категория	36 1/4 - {1.0000000, 2.0000000}	37 2/4 - {2.0000000, 4.0000000}	38 3/4 - {4.0000000, 4.0000000}	39 4/4 - {4.0000000, 5.0000000}
4 4/4 - {59250.0000000, 100860.0000000}				
3 3/4 - {50400.0000000, 59250.0000000}				
2 2/4 - {47310.0000000, 50400.0000000}				
1 1/4 - {43200.0000000, 47310.0000000}

Рисунок 20. Визуализация когнитивных функций для обобщенных классов и всех описательных шкал в модели INF5

2.3. SWOT и PEST матрицы и диаграммы

SWOT-анализ является широко известным и общепризнанным методом стратегического планирования. Однако это не мешает тому, что он подвергается критике, часто вполне справедливой, обоснованной и хорошо аргументированной. В результате критического рассмотрения SWOT-анализа выявлено довольно много его слабых сторон (недостатков), источником которых является необходимость привлечения экспертов, в частности для оценки силы и направления влияния факторов. Ясно, что эксперты это делают неформализуемым путем (интуитивно), на основе своего профессионального опыта и компетенции. Но возможности экспертов имеют свои ограничения и часто по различным причинам они не могут и не хотят это сделать. Таким образом, возникает проблема проведения SWOT-анализа без привлечения экспертов. Эта проблема может решаться путем автоматизации функций экспертов, т.е. путем измерения силы и направления влияния факторов непосредственно на основе эмпирических данных. Подобная технология разработана давно, ей уже около 30 лет, но она малоизвестна – это интеллектуальная система «Эйдос». Данная система всегда обеспечивала возможность проведения количественного автоматизированного SWOT-анализа без использования экспертных оценок непосредственно на основе эмпирических данных. Результаты SWOT-анализа выводились в форме информационных портретов. В версии системы под MS Windows: «Эйдос-X++» предложено автоматизированное количественное решение прямой и обратной задач SWOT-анализа с построением традиционных SWOT-матриц и диаграмм (рисунок 21).

Рисунок 21. Пример SWOT-матрицы в модели INF5

На рисунке 22 приведены примеры инвертированной SWOT-матрицы и инвертированной SWOT-диаграммы в модели INF5.

Рисунок 22. Пример SWOT-матрицы в модели INF5

2.4. Кластерно-конструктивный анализ признаков

На рисунке 23 приведены результаты кластерно-конструктивного анализа признаков:

Рисунок 23. Результаты кластерно-конструктивного анализа признаков

ЗАКЛЮЧЕНИЕ

Так как существует множество систем искусственного интеллекта, то возникает необходимость сопоставимой оценки качества их математических моделей. Одним из вариантов решения этой задачи является тестирование различных системы на общей базе исходных данных, для чего очень удобно использовать общедоступную базу репозитория UCI. В данной работе приводится развернутый пример использования базы данных репозитория UCI для оценки качества математических моделей, применяемых в АСК-анализе и его программном инструментарии системе искусственного интеллекта «Эйдос». При этом наиболее достоверной в данном приложении оказались модели INF4, основанная на семантической мере целесообразности информации А.Харкевича при интегральном критерии «Сумма знаний». Точность модели составляет 0,960, что заметно выше, чем достоверность экспертных оценок, которая считается равной около 70%. Для оценки достоверности моделей в АСК-анализе и системе «Эйдос» используется F-критерий Ван Ризбергера и его нечеткое мультиклассовое обобщение, предложенное проф.Е.В.Луценко. Также обращает на себя внимание, что статистические модели в данном приложении дают примерно на 21% более низкую средневзвешенную достоверность идентификации и не идентификации, чем модели знаний, что, как правило, наблюдается и в других приложениях. Этим и оправдано применение моделей знаний.

На основе базы данных UCI, рассмотренной в данной работе, построить модели прогнозирования не с помощью АСК-анализа и реализующей его системы «Эйдос», а с применением других математических методов и реализующих их программных систем, то можно сопоставимо сравнить их качество.

Литература

1. Луценко Е.В. Методика использования репозитория UCI для оценки качества математических моделей систем искусственного интеллекта / Е.В. Луценко // Политематический сетевой электронный научный журнал Кубанского государственного аграрного университета (Научный журнал КубГАУ) [Электронный ресурс]. – Краснодар: КубГАУ, 2003. – №02(002). С. 120 – 145. – IDA [article ID]: 0020302012. – Режим доступа: <http://ej.kubagro.ru/2003/02/pdf/12.pdf>, 1,625 у.п.л.
2. Луценко Е.В. АСК-анализ, моделирование и идентификация живых существ на основе их фенотипических признаков / Е.В. Луценко, Ю.Н. Пенкина // Политематический сетевой электронный научный журнал Кубанского государственного аграрного университета (Научный журнал КубГАУ) [Электронный ресурс]. – Краснодар: Куб- ГАУ, 2014. – №06(100). С. 1346 – 1395. – IDA [article ID]: 1001406090. – Режим доступа: <http://ej.kubagro.ru/2014/06/pdf/90.pdf>, 3,125 у.п.л.
3. Луценко Е.В. Теоретические основы, технология и инструментарий автоматизированного системно-когнитивного анализа и возможности его применения для сопоставимой оценки эффективности вузов / Е.В. Луценко, В.Е. Коржаков // Политематический сетевой электронный научный журнал Кубанского государственного аграрного университета (Научный журнал КубГАУ) [Электронный ресурс]. – Краснодар: Куб- ГАУ, 2013. – №04(088). С. 340 – 359. – IDA [article ID]: 0881304022. – Режим доступа: <http://ej.kubagro.ru/2013/04/pdf/22.pdf>, 1,25 у.п.л.
4. 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.
5. [Электронный ресурс]. Статья "baza dannix sotrudnikov": <http://allexcel.ru/gotovye-tablitsy-excel-besplatno>, свободный. - Загл. с экрана. Яз. англ.
6. Сайт профессора Е.В.Луценко [Электронный ресурс]. Режим доступа: <http://lc.kubagro.ru/>, свободный. - Загл. с экрана. Яз. рус.
7. Луценко Е.В. 30 лет системе «Эйдос» – одной из старейших отечественных универсальных систем искусственного интеллекта, широко применяемых и развивающихся и в настоящее время / Е.В. Луценко // Политематический сетевой электронный научный журнал Кубанского государственного аграрного университета (Научный журнал КубГАУ) [Электронный ресурс]. – Краснодар: КубГАУ, 2009. – №10(054). С. 48 – 77. – Шифр Информрегистра: 0420900012\0110, IDA [article ID]: 0540910004. – Режим доступа: <http://ej.kubagro.ru/2009/10/pdf/04.pdf>, 1,875 у.п.л.
8. Луценко Е.В. Универсальная когнитивная аналитическая система «Эйдос-Х++» / Е.В. Луценко // Политематический сетевой электронный научный журнал Кубанского государственного аграрного университета (Научный журнал КубГАУ) [Электронный ресурс]. – Краснодар: КубГАУ, 2012. – №09(083). С. 328 – 356. – IDA [article ID]: 0831209025. – Режим доступа: <http://ej.kubagro.ru/2012/09/pdf/25.pdf>, 1,812 у.п.л.
9. Луценко Е.В. Прогнозирование количества и классов солнечных вспышек на основе их предыстории по данным репозитория UCI с применением АСК-анализа и интеллектуальной системы «Эйдос» / Е.В. Луценко, А.Ю. Боровко // Политематический сетевой электронный научный журнал Кубанского государственного аграрного университета (Научный журнал КубГАУ) [Электронный ресурс]. – Краснодар: КубГАУ, 2014. – №10(104). С. 1309 – 1370. – IDA [article ID]: 1041410099. – Режим доступа: <http://ej.kubagro.ru/2014/10/pdf/99.pdf>, 3,875 у.п.л.

10. Луценко Е.В. Моделирование сложных многофакторных нелинейных объектов управления на основе фрагментированных зашумленных эмпирических данных большой размерности в системно-когнитивном анализе и интеллектуальной системе «Эйдос-Х++» / Е.В. Луценко, В.Е. Коржаков // Политематический сетевой электронный научный журнал Кубанского государственного аграрного университета (Научный журнал КубГАУ) [Электронный ресурс]. – Краснодар: КубГАУ, 2013. – №07(091). С. 164 – 188. – IDA [article ID]: 0911307012. – Режим доступа: <http://ej.kubagro.ru/2013/07/pdf/12.pdf>, 1,562 у.п.л.

11. Луценко Е.В. Нечеткое мультиклассовое обобщение классической F-меры достоверности моделей Ван Ризбергена в АСК-анализе и системе «Эйдос» / Луценко Е.В. // Политематический сетевой электронный научный журнал Кубанского государственного аграрного университета (Научный журнал КубГАУ) [Электронный ресурс]. – Краснодар: КубГАУ, 2016. – №09(123). – Режим доступа: <http://ej.kubagro.ru/2016/09/pdf/01.pdf>, 1,813 у.п.л. – IDA [article ID]: 1231609001. <http://dx.doi.org/10.21515/1990-4665-123-001>

12. Луценко Е.В. Количественный автоматизированный SWOT- и PEST-анализ средствами АСК-анализа и интеллектуальной системы «Эйдос-Х++» / Е.В. Луценко // Политематический сетевой электронный научный журнал Кубанского государственного аграрного университета (Научный журнал КубГАУ) [Электронный ресурс]. – Краснодар: КубГАУ, 2014. – №07(101). С. 1367 – 1409. – IDA [article ID]: 1011407090. – Режим доступа: <http://ej.kubagro.ru/2014/07/pdf/90.pdf>, 2,688 у.п.л.

References

1. Lucenko E.V. Metodika ispol'zovaniya repozitorija UCI dlja ocenki kachestva matematicheskikh modelej sistem iskusstvennogo intellekta / E.V. Lucenko // Politematicheskij setevoy jelektronnyj nauchnyj zhurnal Kubanskogo gosudarstvennogo agrarnogo universiteta (Nauchnyj zhurnal KubGAU) [Jelektronnyj resurs]. – Krasnodar: KubGAU, 2003. – №02(002). S. 120 – 145. – IDA [article ID]: 0020302012. – Rezhim dostupa: <http://ej.kubagro.ru/2003/02/pdf/12.pdf>, 1,625 u.p.l.

2. Lucenko E.V. ASK-analiz, modelirovanie i identifikacija zhivyyh sushhestv na osnove ih fenotipicheskikh priznakov / E.V. Lucenko, Ju.N. Penkina // Politematicheskij setevoy jelektronnyj nauchnyj zhurnal Kubanskogo gosudarstvennogo agrarnogo universiteta (Nauchnyj zhurnal KubGAU) [Jelektronnyj resurs]. – Krasnodar: Kub- GAU, 2014. – №06(100). S. 1346 – 1395. – IDA [article ID]: 1001406090. – Rezhim dostupa: <http://ej.kubagro.ru/2014/06/pdf/90.pdf>, 3,125 u.p.l.

3. Lucenko E.V. Teoreticheskie osnovy, tehnologija i instrumentarij avtomatizirovannogo sistemno-kognitivnogo analiza i vozmozhnosti ego primeneniya dlja sopostavimoy ocenki jeffektivnosti vuzov / E.V. Lucenko, V.E. Korzhakov // Politematicheskij setevoy jelektronnyj nauchnyj zhurnal Kubanskogo gosudarstvennogo agrarno- go universiteta (Nauchnyj zhurnal KubGAU) [Jelektronnyj resurs]. – Krasnodar: Kub- GAU, 2013. – №04(088). S. 340 – 359. – IDA [article ID]: 0881304022. – Rezhim dostupa: <http://ej.kubagro.ru/2013/04/pdf/22.pdf>, 1,25 u.p.l.

4. Lucenko E.V. Avtomatizirovannyj sistemno-kognitivnyj analiz v upravlenii aktivnymi ob#ektami (sistemnaja teorija informacii i ee primenenie v issledovanii jekonomicheskikh, social'no-psihologicheskikh, tehnologicheskikh i organizacionno-tehnicheskikh sistem): Monografija (nauchnoe izdanie). – Krasnodar: KubGAU. 2002. – 605 s.

5. [Jelektronnyj resurs]. Stat'ja "baza dannix sotrudnikov": <http://allexcel.ru/gotovyetablelitsy-excel-besplatno,svobodnyj>. - Zagl. s jekrana. Jaz. ang.

6. Sajt professora E.V.Lucenko [Jelektronnyj resurs]. Rezhim dostupa: <http://lc.kubagro.ru/>, svobodnyj. - Zagl. s jekrana. Jaz. rus.

7. Lucenko E.V. 30 let sisteme «Jejdos» – odnoj iz starejsih otechestvennyh universal'nyh sistem iskusstvennogo intellekta, shiroko primenjaemyh i razvivajushhhsja i v nastojashhee vremja / E.V. Lucenko // Politematicheskij setevoj jelektronnyj nauchnyj zhurnal Kubanskogo gosudarstvennogo agrarnogo universiteta (Nauchnyj zhurnal KubGAU) [Jelektronnyj resurs]. – Krasnodar: KubGAU, 2009. – №10(054). S. 48 – 77. – Shifr Informregistra: 0420900012\0110, IDA [article ID]: 0540910004. – Rezhim dostupa: <http://ej.kubagro.ru/2009/10/pdf/04.pdf>, 1,875 u.p.l.

8. Lucenko E.V. Universal'naja kognitivnaja analiticheskaja sistema «Jejdos- H++» / E.V. Lucenko // Politematicheskij setevoj jelektronnyj nauchnyj zhurnal Kubanskogo gosudarstvennogo agrarnogo universiteta (Nauchnyj zhurnal KubGAU) [Jelektronnyj resurs]. – Krasnodar: KubGAU, 2012. – №09(083). S. 328 – 356. – IDA [article ID]: 0831209025. – Rezhim dostupa: <http://ej.kubagro.ru/2012/09/pdf/25.pdf>, 1,812 u.p.l.

9. Lucenko E.V. Prognozirovanie kolichestva i klassov solnechnyh vspyshek na osnove ih predystorii po dannym repozitorija UCI s primeneniem ASK-analiza i intellektual'noj sistemy «Jejdos» / E.V. Lucenko, A.Ju. Borovko // Politematicheskij setevoj jelektronnyj nauchnyj zhurnal Kubanskogo gosudarstvennogo agrarnogo universiteta (Nauchnyj zhurnal KubGAU) [Jelektronnyj resurs]. – Krasnodar: KubGAU, 2014. – №10(104). S. 1309 – 1370. – IDA [article ID]: 1041410099. – Rezhim dostupa: <http://ej.kubagro.ru/2014/10/pdf/99.pdf>, 3,875 u.p.l.

10. Lucenko E.V. Modelirovanie slozhnyh mnogofaktornyh nelinejnyh ob#ektov upravlenija na osnove fragmentirovannyh zashumlennyh jempiricheskikh dannyh bol'shoj razmernosti v sistemno-kognitivnom analize i intellektual'noj sisteme «Jejdos-H++» / E.V. Lucenko, V.E. Korzhakov // Politematicheskij setevoj jelektronnyj nauchnyj zhurnal Kubanskogo gosudarstvennogo agrarnogo universiteta (Nauchnyj zhurnal KubGAU) [Jelektronnyj resurs]. – Krasnodar: KubGAU, 2013. – №07(091). S. 164 – 188. – IDA [article ID]: 0911307012. – Rezhim dostupa: <http://ej.kubagro.ru/2013/07/pdf/12.pdf>, 1,562 u.p.l.

11. Lucenko E.V. Nechetkoe mul'tiklassovoe obobshhenie klassicheskoy F-mery dostovernosti modelej Van Rizbergena v ASK-analize i sisteme «Jejdos» / Lucenko E.V. // Politematicheskij setevoj jelektronnyj nauchnyj zhurnal Kubanskogo gosudarstvennogo agrarnogo universiteta (Nauchnyj zhurnal KubGAU) [Jelektronnyj resurs]. – Krasnodar: KubGAU, 2016. – №09(123). – Rezhim dostupa: <http://ej.kubagro.ru/2016/09/pdf/01.pdf>, 1,813 u.p.l. – IDA [article ID]: 1231609001. <http://dx.doi.org/10.21515/1990-4665-123-001>

12. Lucenko E.V. Kolichestvennyj avtomatizirovannyj SWOT- i PEST-analiz sredstvami ASK-analiza i intellektual'noj sistemy «Jejdos-H++» / E.V. Lucenko // Politematicheskij setevoj jelektronnyj nauchnyj zhurnal Kubanskogo gosudarstvennogo agrarnogo universiteta (Nauchnyj zhurnal KubGAU) [Jelektronnyj resurs]. – Krasnodar: KubGAU, 2014. – №07(101). S. 1367 – 1409. – IDA [article ID]: 1011407090. – Rezhim dostupa: <http://ej.kubagro.ru/2014/07/pdf/90.pdf>, 2,688 u.p.l.