

сколько локальных экстремумов и другие методы (в частности градиентный метод) могут остановиться на одном из них.

Для решения описанных выше задач РСПЗ/ОУ была разработана программа на языке C++ с использованием стандартных библиотек языка C++ и выполнены расчеты для ряда структур РВС.

ЛИТЕРАТУРА

[1]. Турута Е.Н. Организация распределения задач в вычислительных системах, обеспечивающая их отказоустойчивость// Автоматика и вычисл. техника. 1985. № 1. С.5-14.

- [2]. Tourouta E.N. The methods for ensuring fault-tolerance of distributed control systems// IFAC Symp. on Fault Detection, Supervision and Safety for Technical Processes (SAFEPROCESS' 94). 1994, Helsinki, Finland. Preprints, v. 2, pp.772-777.
- [3]. Корнеев В.В., Гарев А.Ф., Васютин С.В., Райх В.В. Базы данных. Интеллектуальная обработка информации. М.: "Нолидж", 2000.
- [4]. Goldberg D.E. Genetic algorithms in search, optimization and machine learning. N.Y. Addison-Wesley Pub. Comp., 1989.

ИСПОЛЬЗОВАНИЕ ЭКСПЕРТНЫХ СИСТЕМ ДЛЯ ЗАЩИТЫ ИНФОРМАЦИОННО-ВЫЧИСЛИТЕЛЬНЫХ СЕТЕЙ ОТ НЕСАНКЦИОНИРОВАННОГО ДОСТУПА

А.М. Курносков

АННОТАЦИЯ

Рассматриваются вопросы использования экспертных систем для классификации информационного оружия, применяемого в телекоммуникационных сетях. Анализируются существующие методы задания классов состояний объекта, а также алгоритмы кластер-анализа. Алгоритм нечеткой кластеризации Данна-Бежdeka выбран, как наиболее оптимальный, позволяющий системе функционировать в условиях неопределенности.

В настоящее время широкое внедрение телекоммуникационных систем, прежде всего автоматизированных систем управления (АСУ) бизнес-процессами, становится определяющим фактором эффективности любой экономической структуры. Вместе с тем, как показывает мировая практика, растет потенциальная уязвимость экономики по отношению к информационным воздействиям.

АСУ бизнес-процессами представляет собой человеко-машинную систему, использующую современные экономико-математические методы, автоматические средства обработки данных (ЭВМ, автоматических устройств накопления, регистрации, отображения) и связи, а также новые организационные принципы управления. АСУ предназначена для принятия оптимальных решений по управлению производственно-хозяйственной деятельностью корпорации.

Одним из основных элементов системы управления, является информационно-

вычислительная сеть (ИВС), объединяющая следующие элементы:

- сеть обмена данными;
- локальные вычислительные сети (ЛВС) и АСУ региональных представительств, подключенных к базовой сети;
- подсети рабочих групп в рамках подразделений.

К особенностям функционирования перспективной ИВС можно отнести:

- разнородность технических средств и программного обеспечения;
- высокая доля разнородных сетевых информационных технологий;
- расширение спектра услуг и, как следствие, увеличение объема и качественного разнообразия информации, хранимой, обрабатываемой и передаваемой в ИВС;
- увеличение числа пользователей ИВС.

Для перспективной ИВС любой корпорации может стать актуальным следующее:

- сопряжение с глобальными сетями типа Internet;
- сопряжение с сетями других корпораций в результате расширения международного сотрудничества.

Перечисленные особенности корпоративной ИВС, обуславливают высокую степень уязвимости информации, хранимой, обрабатываемой и передаваемой в такой сети.

Очевидно, что при создании единой ИВС корпорации в ее системе защиты информации

сразу не удастся учесть все способы и возможные каналы информационного воздействия на сеть. Учитывая то, что во многих странах глубоко осознали важность информационных ресурсов в ходе как экономического, так и военного противоборства, и значительное внимание уделяется вопросам ведения информационных войн, разрабатываются новые виды информационного оружия, в ИВС должны функционировать эффективные механизмы противодействия информационным воздействиям со стороны противника на сеть.

Рассматривая виды ответных действий средств защиты информации (СЗИ) по отношению к противнику в ИВС можно выделить два направления: простое реагирование и интеллектуальное противодействие (ИП).

Под интеллектуальным противодействием – будем понимать комплекс решаемых в ИВС задач по реагированию на несанкционированные действия (НСД) на основе оперативного анализа стратегии противника и применяемого им информационного оружия, технических возможностей ИВС, текущих задач информационной борьбы и управления, в том числе, и с использованием экспертных систем (ЭС) [1].

Интеллектуальное противодействие подчинено следующим целям информационной борьбы в ИВС:

- снизить время бесконтрольного присутствия противника в ИВС;
- дезинформировать противника;
- дезорганизовать действия противника по осуществлению информационных атак в ИВС;
- снизить не целевую нагрузку на ИВС;
- воздействовать на ресурсы противника.
- Основные задачи ИП следующие:
- установление факта НСД;
- установления характера НСД;
- выбор стратегии ответных действий.

Чтобы принять решение об ответных действиях необходимо, прежде всего, классифицировать НСД.

В общем случае классификация НСД включает в себя:

- определение субъекта НСД;
- определение объекта НСД;
- классификацию используемого субъектом информационного оружия.

Рассмотрим некоторые методы классификации состояний объекта в экспертных системах. Всякому реальному объекту можно поставить в соответствие его состояние. Пусть состояние

объекта или предметной области можно охарактеризовать значениями некоторых признаков или параметров. Если множество состояний объекта обладает общими свойствами, или значениями признаков описания состояний объекта по их схожести, то процесс классификации можно представить в виде двух этапов:

- изучение и установление общих свойств группы состояний одного и того же класса, или выделение класса;
- принятие решения, устанавливающего принадлежность нового, ранее в процессе классификации не рассматриваемого, входного состояния к данному классу путем сравнения значений признаков (свойств) этого состояния с уже известным состоянием, присущим данному классу.

Те состояния, для которых заранее известна их принадлежность к тому или иному классу, называют обучающими состояниями, или образцами. Те, которые требуется классифицировать – испытуемыми состояниями.

Другими словами, экспертная система, выполняющая классификацию состояний, производит обнаружение и выделение обоих признаков состояний, принадлежащих к одному классу, а также вводимого в ЭС состояния к одному из заданных классов.

Существует несколько методов задания классов состояний объекта [2]:

- метод перечисления членов класса;
- метод общности свойств;
- метод кластеризации;

Задание класса перечислением состояний, входящим в его состав, заключается в реализации процедуры сравнения с эталоном. Состояния, принадлежащие к одному и тому же классу, запоминаются ЭС. При предъявлении новых состояний ЭС последовательно сравнивает их с хранящимися состояниями. Новое состояние относится к такому классу, к которому принадлежало состояние, совпавшее с новым. Однако в тех случаях, когда классы состоят из большого числа элементов, и каждый элемент описывается значениями большого числа признаков, реализация этого принципа сильно затруднена.

Задание класса при помощи свойств, общих для всех состояний данного класса, заключается в выделении признаков описания состояний объекта и работы с ними. Состояния, принадлежащие одному и тому же классу, обладают рядом общих свойств, или признаков, отражающих подобие этих состояний. Когда ЭС предъявляется неклассифицированное состояние, выделяется набор описывающих его при-

знаков, после чего значения этих признаков сравниваются с набором значений признаков, соответствующих каждому классу. ЭС причисляет состояние к тому классу, который имеет набор значений признаков классифицирующего состояния.

Поскольку значения признаков, характеризующих класс, находятся в некотором диапазоне, метод сопоставления признаков допускает некоторое отличие значений признаков отдельных состояний, чего не допускает процедура сравнения с эталоном. Исключительно трудно, если не невозможно вообще найти для некоторого класса полный набор различающих признаков, что делает применение данного метода достаточно сложным.

Когда образ некоторого класса можно представить как вектор чисел, определяющего геометрическое расположение состояний в пространстве, координатами которого являются признаки описания объекта, говорят о кластеризации.

Группа состояний объекта, образующих в пространстве описаний компактную в некотором смысле область, называется кластером. Реализация метода кластеризации в ЭС определяется взаимным пространственным расположением кластеров в пространстве. Если кластеры, соответствующие разным классам, разнесены достаточно далеко друг от друга, можно воспользоваться классификацией по какой-либо из метрик. Если кластеры перекрываются, что происходит в результате неполной информации о состоянии объекта, применяются более сложные методы разбиения пространства состояний. Таким образом, наиболее приемлемым методом задания классов состояний объекта по отношению к другим формам представления является метод кластеризации.

Рассмотрим методы кластер-анализа. Пусть состояние объекта имеет размерность N и представляет собой точку в N -мерном пространстве. Каждая координата в пространстве соответствует одному из признаков описания состояния объекта.

Одной из основных задач является определение многомерных областей в пространстве состояний, в которых содержатся состояния, принадлежащие к различным классам. Пространство описаний состояний может быть преобразовано из N -мерного в пространство меньшей размерности с определенной потерей информации. Этот переход осуществляется при помощи взвешивания каждого признака. Для этого всем используемым признакам приписы-

вается вес, больший нуля, а неиспользуемым – вес, равный нулю.

Меры сходства, используемые в кластерном анализе, как правило, имеют вид [3]:

$$d(S_i, S_j) = (1/n \sum_{k=1}^n |a_{ik} - a_{jk}|^r)^{1/r}$$

где

S_i, S_j – возможные состояния объекта;

a_{ik}, a_{jk} – значения признака описания состояний S_i, S_j ;

n – число признаков;

r – положительное целое число.

Расстояние является функцией, вполне отвечающей интуитивному пониманию экспертом сходства объектов. Тем не менее, определение подобия через расстояние между состояниями не является единственно возможным. Главное заключается в том, чтобы подобрать такую меру, при помощи которой можно принять правильное (с точки зрения эксперта) решение о принадлежности или не принадлежности состояния некоторому классу.

В качестве меры сходства, применяемой в кластер-анализе, может быть использован коэффициент общего сходства Гауэра:

$$S_{jk} = \sum_{i=1}^n W_{ijk} S_{ijk} / \sum_{i=1}^n W_{ijk}$$

где

$0 \leq S_{ijk} \leq 1$ – значение сходства между значениями признака i для состояний j и k ,

W_{ijk} – вес, приписываемый признаку;

n – число признаков.

Среди параметров, используемых в различных методах кластерного анализа, можно выделить следующие: плотность состояний в пространстве состояний в пространстве описаний состояний объекта, объем, занимаемый кластером, связность элементов определенного кластера, промежутки между соседними кластерами в сравнении с их диаметрами и другие.

Критерием оптимальности произведенной классификации состояний может служить функция качества:

$$\sum_k \sum_t \sum_n (x_{ij} - \overline{x_{iM}})^2$$

где

k – число кластеров;

t_m – число состояний в кластере M ;

n – число признаков;

x_{ij} – значение i -го признака j -го состояния;

$\bar{x}_{iM}$ – среднее значение i -го признака для кластера M .

Например, объединением кластеров, дающих наименьший вклад в получаемое значение качества, достигается оптимальность данного критерия для сгруппированных таким образом кластеров (алгоритм Уорда).

Рассмотрим существующие алгоритмы формирования кластеров «без учителя» [2]. Целью всех этих алгоритмов помимо формирования кластеров является нахождение для каждого кластера так называемого центра, являющегося его самым представительным эталоном.

При описании алгоритмов в качестве меры сходства будет использовано евклидово расстояние. Одним из самых простых алгоритмов кластеризации является следующий.

Пусть задано множество состояний $S_1, S_2, S_3, \dots, S_n$. Центром первого кластера принимается любое состояние S_1 из заданного множества. Пороговой величиной, относительно которой можно судить о принадлежности состояния к тому или иному кластеру, является положительное число T из интервала $(0; 1)$.

Для простоты полагаем центром первого кластера состояние S_1 . Вычислим степень сходства между центром первого кластера S_1 и состоянием S_2 . Если полученное значение меньше T (в том случае, когда в качестве меры сходства используется мера расстояния, иначе – больше), то состояние S_2 включается в первый кластер. Иначе, состояние S_2 становится центром нового, второго, кластера. Такая же процедура производится и для остальных состояний из заданного множества. Если значение меры сходства очередного состояния S_i и каждого определенного ранее центра кластера превышает значение T , S_i становится центром нового кластера. Иначе, состояние S_i включается в состав кластера с наиболее сходным к состоянию S_i центром.

Недостатком предложенного алгоритма является зависимость получаемой классификации состояний от выбора первого центра кластера, а также от значения порога T .

Алгоритм максиминного расстояния. На первом шаге алгоритма в качестве центра первого кластера произвольным образом выбирается любое состояние S_j из заданного множества состояний. После этого отыскивается состояние, находящееся на наибольшем для данного центра расстоянии, которое и назначается центром второго кластера. На последующих шагах вычисляются расстояния между осталь-

ными состояниями и уже определенными центрами кластеров. Среди вычисленных расстояний выбирается минимальное. Затем определяется максимальное расстояние среди всех минимальных. Если последнее составляет значительную часть расстояния между центрами кластеров (скажем, по меньшей мере половину этого расстояния), соответствующее состояние назначается центром очередного кластера. Иначе все кластеры определены.

Оставшиеся состояния классифицируются следующим образом. Состояние S_k заносится в тот кластер, расстояние, с центром которого является минимальным. Производится уточнение каждого центра кластера посредством определения для каждого центра кластера выборочного среднего по формуле:

$$a_i^j = 1/N \sum_{k=1}^N a_{ik}, i = \overline{1, n}$$

где

j – номер кластера;

N_j – число выявленных состояний j -го кластера;

a_{ik} – значение i -го признака k -го состояния.

Точки в пространстве описаний состояний с координатами $(a_1^1, a_2^1, \dots, a_n^1)$ принимаются в качестве новых центров кластеров.

Алгоритм K внутригрупповых средних. Задается множество кластеров – K , которое определяется в соответствии с особенностями объекта, состояние которого классифицируется. Произвольным образом в качестве центров кластеров выбираются состояния из предлагаемого для классификации множества состояний. На первом шаге множество состояний разбивается на кластеры следующим образом. Состояние заносится в тот кластер, расстояние, с центром которого минимально. Если расстояния с несколькими центрами одинаковы, состояния классифицируются в один из этих кластеров произвольным образом. После произведенной классификации всех состояний производится коррекция центров кластеров по формуле:

$$z_{ij}^{(l+1)} = 1/N \sum_{m=1}^N a_{mi}, i = \overline{1, n}, j = \overline{1, k} \quad (1)$$

где

$z_{ij}^{(l+1)}$ – i -тая координата центра j -го кластера на следующем, $l+1$, шаге;

N_j – количество состояний, относящихся к j -му кластеру на l -том шаге;

a_{mi} – i -тая координата m -го состояний из множества состояний, относящихся к j -му кластеру.

Если $z_{ij}(l+1) = z_{ij}(l)$, $\forall i = 1, \dots, n, j = 1, \dots, k$, где

n – количество признаков описания состояний объекта, то алгоритм завершен. Иначе состояния классифицируются по кластерам с новыми центрами, после чего опять производится коррекция центров кластеров.

Результаты, получаемые при выполнении данного алгоритма, во многом зависят от числа K и выбора центров кластеров на первом шаге.

Алгоритм Isodata. Рассмотрим пошаговое выполнение данного алгоритма. До выполнения алгоритма следует задать набор N_c исходных центров кластеров.

Шаг 1. Пусть K – необходимое число кластеров; L – максимальное число пар кластеров, которых можно определить за одну итерацию; Θ_c – параметр, характеризующий компактность пространственного расположения кластеров; Θ_s – параметр, характеризующий среднеквадратичное отклонение; Θ_N – параметр, определяющий количество состояний, находящихся в кластере.

Из выборки предлагаемых для классификации состояний $S_1, S_2, \dots, S_N$ произвольным образом выбираются центры кластеров $z_1, z_2, \dots, z_{N_c}$, где N_c необязательно совпадает с K .

Шаг 2. Состояния $S_1, S_2, \dots, S_N$ заносятся в кластеры следующим образом. Состояние классифицируется в тот кластер, расстояние с которым минимально. Если расстояния с несколькими центрами кластеров одинаковы, состояние заносится в один из этих кластеров произвольным образом. Получается множество состояний $S_j, j = 1, \dots, N_c$.

Шаг 3. Если мощность множества S_j для какой либо j меньше Θ_N , центр кластера удаляется из множества центров кластеров, значение N_c уменьшается на единицу и множество S_j дальше не рассматривается.

Шаг 4. Производится коррекция центров по формуле (1).

Шаг 5. Вычисляются средние внутрикластерные расстояния D_j по формуле:

$$\overline{D}_j = 1/N_j \sum_{x \in S_j} \sqrt{\sum_{i=1}^n (x_i - z_{ji})^2}, j = \overline{1, N_c}$$

где

z_{ij} – i -тая координата центра j -го кластера.

Шаг 6. Определяется обобщенное среднее расстояние по формуле:

$$\overline{D} = 1/N \sum_{j=1}^N N_j D_j$$

Шаг 7. Проверяются следующие условия:

- если текущая итерация последняя, то $\Theta_c = 0$ и производится переход к шагу 11;
- если $N_c \leq k/2$, то производится переход к следующему, восьмому шагу;
- если текущая итерация имеет четкий порядковый номер или выполняется соотношение $N_c \geq 2k$, то производится переход к шагу 11.

Шаг 8. Вычисляются среднеквадратичные отклонения по всем координатам для каждого кластера по формуле:

$$\sigma_{ij} = \sqrt{1/N_j \sum_{x \in S_j} (x_{ik} - z_{ij})^2}, i = \overline{1, n}, j = \overline{1, N_c}$$

Шаг 9. Определяется $\sigma_{j \max} = \max \sigma_{ij}, i = \overline{1, n}, j = \overline{1, N_c}$.

Шаг 10. Если для какого-либо j выполняется состояние $\sigma_{j \max} > \Theta_s$, а также выполняется соотношение $N_c \leq K/2$ или соотношение $D_j > D$ и $N_j > 2(\Theta + 1)$, то j -тый кластер разделяется на два кластера с центрами z_j^+ и z_j^- , N_c увеличивается на единицу, центр z_j удаляется из множества центров кластеров. Координата центров z_j^+ и z_j^- , соответствующая $\sigma_{j \min}$, увеличивается и уменьшается соответственно на некоторую величину φ_j ($\varphi_j = k\sigma_{j \max}$, $0 < k \leq 1$).

Если хотя бы один кластер был разделен, то переходят к шагу 2.

Шаг 11. Определяется расстояние между всеми парами центров кластеров по формуле:

$$D_{ij} = \sqrt{\sum (z_{ik} - z_{jk})^2}, i = \overline{1, N_c - 1}, j = \overline{i + 1, N_c}$$

Шаг 12. L расстояний D_{ij} , которые меньше Θ_c , упорядочиваются по возрастанию.

Шаг 13. Слияние кластеров. В последовательности, определенной на предыдущем шаге, кластеры с центрами z_{i1} и z_{j1} , $l = \overline{1, L}$, объединяются при условии, что на текущем шаге слияние не применялось ни к одному из этих двух кластеров. Координаты нового центра вычисляются по формуле:

$$z_{kl}^* = 1/(N_{i1} + N_{j1}) [N_{i1}(z_{i1}) + N_{j1}(z_{j1})]$$

N_c уменьшается на единицу, центры z_{i1} и z_{j1} удаляются из множества центров кластеров.

Шаг 14. Текущая итерация завершена. Если она не является последней, то переход к шагу 1, если необходимо изменить значение какого-либо параметра алгоритма, иначе – к шагу 2.

Если итерация последняя, то алгоритм завершен.

Получение приемлемых (с точки зрения эксперта) результатов классификации по данному алгоритму возможно лишь после анализа большого числа экстремумов, связанных с изменением параметров алгоритма на основании полученных при выполнении алгоритма харак-

теристик. Таких, как, например, расстояние между центрами кластеров, среднеквадратичные внутрикластерные отклонения, ближайшие и наиболее удаленные от центра состояния кластера и другие.

Алгоритм нечеткой кластеризации Данна-Беждека [4]. Пусть X множество состояний некоторого объекта, $\mu_1, \mu_2, \dots, \mu_k$ – функции принадлежности нечетким кластерам $F_1, F_2, \dots, F_k$. Нечеткие кластеры образуют нечеткое покрытие множества X в том и только том случае,

$$J(\mu) = \min_V \sum_{i=1}^k \sum_{x \in X} (\mu_i(x))^2 \|x - V_i\|^2$$

если $\mu_1(x) + \mu_2(x) + \dots + \mu_k(x) \geq 1, \forall x \in X$.

Качество нечеткого покрытия можно оценить при помощи следующей характеристики: где

$V_1, V_2, \dots, V_k$ – центры кластеров,
 $V_i \in L$ – векторное пространство с нормой $\|\cdot\|$, порожденной скалярным произведением.

$J(\mu)$ оценивает среднеквадратичные отклонения состояний из X относительно центров $V_1, V_2, \dots, V_k$.

Шаг 1. По одному из алгоритмов, приведенных выше (например, Isodata), производится начальная классификация состояний по кластерам, причем $\mu_i(x)$, $i = 1, k$, характеризует близость состояния x центра i -го кластера.

Шаг 2. Центры кластеров уточняются при

$$V_i = \sum_{x \in X} (\mu_i(x))^2 x / \sum_{x \in X} (\mu_i(x))^2, \\ i = 1, k, x \in X \subset L$$

помощи формулы:

Шаг 3. Строится новое покрытие $F_1, F_2, \dots, F_k$, описываемое при помощи $\mu_1, \mu_2, \dots, \mu_k$ в соответствии с правилом:

$$\hat{\mu}_i(X) = \frac{1}{\sum_{j=1}^k \left(\frac{1}{\|X - V_j\|^2} \right)}$$

Шаг 4. Вычисляются отклонения $\delta \leq \varepsilon$ – некоторого порога, то алгоритм завершен. Иначе производится переход к шагу 2.

Размер обучающей выборки, необходимой для построения прикрытия пространства состояний объекта кластерами, относится к числу неопределенных параметров, которые должны быть определены в процессе обучения.

Рассмотрим возможность классификация текущих состояний объекта в ЭС. Пусть в ЭС при помощи одного из алгоритмов кластерного анализа определено K кластеров с центрами $Z_1, Z_2, \dots, Z_k$. Тогда процесс классификации текущего состояния объекта S_0 , подаваемого на вход ЭС, заключается в сопоставлении S_0 с каждым из этих центров и принятии решения о принадлежности S_0 тому кластеру, расстояние с которым минимально.

При этом данное расстояние не должно превышать некоторого порога T , определяемого заранее. Если минимальное расстояние превышает T , то вводится новый кластер с центром, совпадающим с входным состоянием. Это указывает на то, что обучающая выборка, сформированная для задания кластера на этапе обучения, была не полна и не включала состояния, сходные поступившему на вход ЭС состоянию. Можно отметить, что задание порога T может существенным образом повлиять на результаты классификации.

Пусть в процессе кластер-анализа было построено нечеткое покрытие пространства состояний объекта нечеткими кластерами $F_1, F_2, \dots, F_k$ и на вход ЭС подается текущее состояние объекта S_0 . В процессе распознавания S_0 определяется множество значений $\mu_{01}, \mu_{02}, \dots, \mu_{0k}$, характеризующее соответствие состояния S_0 каждому кластеру F_i , $i=1..k$. Если получаемые значения μ_{0i} приводят к выполнению соотношения $\mu_{01} + \mu_{02} + \dots + \mu_{0k} > 1$ или $\max_i \mu_{0i} \geq T$, где i – номер кластера, T – некоторый порог, то принимается решение о создании нового нечеткого кластера F_{k+1} , центром которого является S_0 .

Если в процессе классификации текущего состояния объекта S_0 , был сформирован новый кластер $k+1$ и каждому i – му кластеру ($i=1..k$) экспертом было сформировано управляющее воздействие на объект, то состояние S_0 предъявляется эксперту для определения управляющего воздействия, соответствующего данному состоянию.

Учет неполноты информации о состоянии объекта (ситуации, в которых находится объект) заключается в следующем. Пусть для признаков описания состояния объекта $X_1, X_2, \dots, X_q$ неизвестны их значения в некоторой ситуации S_i . Тогда соответствующий элемент a_k^i , $k=1..q$, характеризующий степень сходства ситуаций S_i и центра кластера Z_j имеет значение 0.5, что означает неопределенность в отношении этого признака при сравнении S_i и Z_j . Чем больше число неопределенных признаков q и

выше вес признаков ω_k , $k=1..q$, тем ниже качество классификации состояний.

Таким образом, для успешного решения поставленной задачи наиболее оптимальным является алгоритм нечеткой классификации Данна-Беждека позволяющий системе работать в условиях неопределенности, что особенно важно, так как с этой проблемой приходится сталкиваться практически в любых ситуациях.

ЛИТЕРАТУРА

- [1].Гриняев С.Н. Интеллектуальное противодействие информационному оружию. – М.: СИНТЕГ, 1999.
- [2].Ту Дж., Гонсалес Р. Принципы распознавания образов. – М.: Мир, 1978. – 415с.
- [3].Сокал Р.Р. Кластер- анализ и классификация: предпосылки и основные направления// Классификация и кластер /Под ред. Дж. Вэн Райзина. – М.: Мир, 1980. С.7-19.
- [4].Заде Л.А. Размытые множества и их применение в распознавании образов и кластер-анализе//Классификация и кластер /Под ред. Дж. Вэн Райзина. – М.: Мир, 1980.

ОБ ОДНОМ ПОДХОДЕ К СОГЛАСОВАНИЮ СРЕДСТВ ИЗВЛЕЧЕНИЯ ЗНАНИЙ

В.В. Смирнов

Московский государственный инженерно-физический институт (технический университет),
Каширское ш., 31, Москва, 115409, РОССИЯ, тел. (095) 324-28-85, Vitaly_Smirnov@mail.ru

АННОТАЦИЯ

В работе рассматривается подход к согласованию средств извлечения знаний из экспертов, проблемно-ориентированных текстов и баз данных, использованный в инструментальном комплексе АТ-ТЕХНОЛОГИЯ. Подход предусматривает как согласование обмена данными между компонентами, так и передачи им управления, что обеспечивается организацией взаимодействия компонент по типу “классной доски”, а также применением следующих средств инструментального комплекса АТ-ТЕХНОЛОГИЯ: лингвистического процессора, интерпретатора сценариев диалога с экспертом, средств верификации поля знаний и интеллектуального планировщика.

1. ВВЕДЕНИЕ

В работе [1] были описаны вопросы автоматизированного построения баз знаний (БЗ) для интегрированных экспертных систем (ИЭС) на основе задачно-ориентированной методологии (ЗОМ) и поддерживающего эту методологию инструментального комплекса АТ-ТЕХНОЛОГИЯ [2]. В работе [3] рассмотрено расширение ЗОМ за счет применения баз данных (БД) как дополнительного источника знаний при формировании БЗ. Автоматизированное построение БЗ средствами комплекса АТ-ТЕХНОЛОГИЯ осуществляется в несколько этапов. В данной работе рассматривается под-

ход к согласованию выделенной группы средств, которая обеспечивает этап извлечения знаний (ИЗ) из экспертов, проблемно-ориентированных текстов и баз данных. Фрагмент информационно-логической (ИЛ) модели данной группы средств показан на Рис. 1.

Реализованные компоненты средств ИЗ относятся к одной из следующих групп:

- компонентам специализированного лингвистического процессора, обеспечивающим определение типа решаемой задачи [4] для активизации сценария диалога с экспертом, сбор лексики системного аналитика [5,6] и эксперта из текста с описанием решаемой задачи [7] и протокола интервьюирования эксперта (ПИЭ) [5] с формированием вспомогательных словарей новой лексики [6], получение информации о наличии НЕ-факторов [6,8];
- компонентам интерпретатора сценариев диалога с экспертом, обеспечивающим активизацию сценариев диалога с экспертом, каждый из которых отражает структуру решаемой задачи и обеспечивает способ ввода информации от эксперта по методу “имитация консультации” [9], для выявления дифференцирующих условий поддерживает метод репертуарных решеток [9] и организует взаимодействие с компонентами, извлечения знаний из данных, описанных в работе [3];