УДК 681.3

СПОСОБ ВЫЯВЛЕНИЯ НЕГАТИВНЫХ ВОЗДЕЙСТВИЙ НА ИНФОРМАЦИОННЫЕ СИСТЕМЫ ЗАЩИТЫ ИНФОРМАЦИИ

С.В. Скрыль, С.В. Белокуров, А.И. Бороненков, П.Е. Краснов

Рассматриваются существующие методы выявления негативных последствий и производится оценка их функционирования в условиях негативных воздействий при наличии жестких ограничений на временной и вычислительный ресурс системы защиты информации. Предложен новый способ выявления негативных воздействий и произведена его оценка

Ключевые слова: распознавание, класс, признак, объект, метод, расстояние, мера близости, граница, информационное пространство

Распознавание, как известно, является одной из ключевых проблем выявления негативных воздействий на информационные системы инновационного характера (ИСИХ), особенно остро стоящей в условиях ограничения временного и вычислительного ресурса на работу системы защиты информации (СЗИ).

Также необходимо отметить, что задача распознавания информационного воздействия сводится к решению трех подзадач: сбор первичной информации; ее предварительная обработка с целью формирования совокупности признаков (нормировка, отбор, сопоставление и т.д.) и классификация, т.е. принятие решения о том, к какому из заданных классов относится наблюдаемое воздействие.

Рассмотрим более конкретно задачу классификации. Эффективность выполнения данной задачи в конкретных условиях напрямую зависит от выбранного метода классификации. Все известные методы, использующие понятие расстояния, предполагают наличие одного или нескольких эталонов для каждого класса.

Задача распознавания может быть сформулирована следующим образом. Имеется множество взаимоисключающих классов $\Omega = (\Omega_1,...,\Omega_m)$, каждый класс состоит из объектов, принадлежность объектов классу определяется соответствующей ему совокупностью признаков, причем количество признаков для всех классов фиксировано. Рассматривается некоторый объект ω, представленный в виде результатов наблюдений (измерений) его признаков. Задача распознавания состоит в том, чтобы отнести исследуемый объект ω к одному из взаимоисключающих классов Ω_i, где $i = \overline{1,m}$.

Условиями выполнения этой задачи является применение системы выявления негативных воздействий на информационную систему, используемую при интенсивном информационном противодействии, как следствие, ограниченность времени для принятия решения о контрпротиводействии. Таким образом, способ распознавания негативных воздействий должен обеспечивать эффективное функционирование информационной системы в динамично меняющихся условиях ведения информационного противоборства. Рассмотрим известные методы распознавания. Существует метод распознавания, основанный на использовании алгоритма минимума расстояния [2-4]. В данном методе используется словарь признаков (таблица).

В нем на языке этих признаков описаны объекты $\omega_{ij} = \left(X_{ij1}, X_{ij2},..., X_{ijN}\right)$, где $i = \overline{1,m}$, $j = \overline{1,r}$, которые составляют классы распознавания Ω_i, где $i = \overline{1,m}$. При появлении в информационном N-мерном пространстве I воздействия $\omega = \left(x_1, x_2,..., x_N\right)$, не совпадающего по

Скрыль Сергей Васильевич - ВИ МВД России, д-р техн. наук, профессор, тел. 8(473) 262-33-76
Белокуров Сергей Владимирович - ВИ ФСИН России д-р техн. наук, профессор, тел. 8(473) 260-68-19
Бороненков Александр Иванович - ВИ МВД России, адъюнкт, тел. 8(473) 262-33-76
Краснов Петр Евгеньевич - ВИ МВД России, аспирант, e-mail: kraspetert@bk.ru

характеристикам ни с одним объектом ни одного из известных классов воздействий, вычисляют среднеквадратичное расстояние L_i, где $i=\overline{1,m}$, между распознаваемым объектом и классом путем перебора и усреднения расстояний d_{ij}^2, где $i=\overline{1,m}$, $j=\overline{1,r}$, между неизвестным объектом и объектами, составляющими класс (рис. 1).

Для этого по формуле (1) рассчитывают расстояние между объектами

$$d^2{}_{ij}=\sum_{p=1}^{N}(x_{\omega_{ij}}^{(p)}-x_{\omega}^{(p)})^2, \qquad (1)$$

где $i=\overline{1,m}$, $j=\overline{1,r}$.

Словарь признаков

Классы	Объекты	Значения признаков			
		X_1	X_2	…	X_N
Ω_1	ω_{11}	$x_{11,1}$	$x_{11,2}$	…	$x_{11,N}$
	ω_{12}	$x_{12,1}$	$x_{12,2}$	…	$x_{12,N}$
	…	…	…	…	…
	ω_{1r}	$x_{1r,1}$	$x_{1r,2}$	…	$x_{1r,N}$
…	…	…	…	…	…
Ω_m	ω_{m1}	$x_{m1,1}$	$x_{m1,2}$	…	$x_{m1,N}$
	ω_{m2}	$x_{m2,1}$	$x_{m2,2}$	…	$x_{m2,N}$
	…	…	…	…	…
	ω_{mr}	$x_{mr,1}$	$x_{mr,2}$	…	$x_{mr,N}$

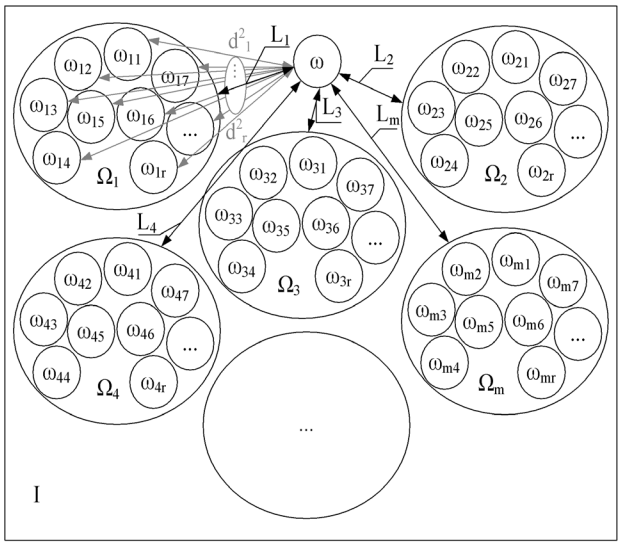

Рис. 1. Определение меры близости с использованием алгоритма минимума расстояний

Далее вычисляют среднеквадратичное расстояние $L(\omega,\Omega_i)$ между распознаваемым объектом и классом по формуле

$$L(\omega,\Omega_i)=\sqrt{\frac{1}{r_i}\sum_{j=1}^{r_i}d^2(\omega,\omega_{ij})}, \qquad (2)$$

где $i=\overline{1,m}$, $j=\overline{1,r}$.

Подобную процедуру повторяют для каждого класса. На основании близости к какому-либо классу принимают решение о принадлежности рассматриваемого объекта ω к этому классу Ω_i, где $i=\overline{1,m}$. Решающее правило выглядит следующим образом:

$$\omega\in\Omega_i, \text{ если } L_i(\omega,\Omega_i)=\min\{L_i(\omega,\Omega_i)\}. \qquad (3)$$

Данный метод обладает следующими недостатками: большие вычислительные затраты при увеличении количества объектов рассматриваемых классов и анализируемых признаков и, как следствие, большое количество времени при принятии решения.

Еще одним методом распознавания является метод, основанный на использовании алгоритма «ближайших соседей» [1]. Сущность этого метода похожа на метод, основанный на использовании алгоритма минимума расстояний, однако, принадлежность неизвестного объекта ω классу Ω_i, где $i=\overline{1,m}$, определяется вычислением меры близости для класса не усреднением расстояний d_{ij}^2, где $i=\overline{1,m}$, $j=\overline{1,r}$, а определением наибольшего количества из K объектов, принадлежащих одному классу и находящихся на минимальном от него расстоянии. Решающее правило выглядит следующим образом:

$$\omega\in\Omega_i, \text{ если } L=\max_i\sum_{b=1}^{K}\delta_{ib}, \qquad (4)$$

где δ_{ib} – символ Кронекера, $\delta_{ib}=1$ при $b\in i$, $i=\overline{1,m}$ и $\delta_{ib}=0$ при $b\notin i$, $i=\overline{1,m}$, K – пороговое значение ближайших соседей.

Недостатками данного метода являются, как и предыдущего, большие вычислительные затраты при увеличении количества рассматриваемых объектов и

анализируемых признаков, а также малая эффективность при осуществлении негативного воздействия.

Также существует метод, предложенный Журавлевым Ю.И., относящийся к многоэтапным методам распознавания и основанный на использовании алгоритма «вычисления оценок» (АВО) [2-4]. Сущность метода заключается в определении степени похожести распознаваемого объекта, в нашем случае вредоносного воздействия, на объекты, входящие в состав априорно известного класса по так называемым системам опорных множеств.

Опыт решения задач распознавания свидетельствует о том, что часто основная информация заключена не в отдельных признаках, а в их сочетаниях. Поскольку не всегда известно, какие именно сочетания информативны, то в методе АВО степень похожести объектов вычисляется не последовательным сопоставлением отдельных признаков, а сопоставлением всех возможных (или определенных) сочетаний признаков, входящих в описание объектов. Совокупность признаков, принятых в качестве информативных в каждом конкретном случае, составляет систему опорных множеств Γ_t, где t изменяется от 1 до максимально принятого значения опорных множеств.

В общем случае реализация метода АВО сводится к формализации следующих этапов:

1) выделяется система опорных множеств Γ_t, по которым производится анализ распознаваемых объектов;

2) вводится понятие близости на множестве частей описаний объектов;

3) задаются правила.

Решающее правило в данном методе может принимать различные формы, в частности распознаваемая строка признаков может быть отнесена к классу, которому соответствует максимальная оценка меры близости, например,

$$\omega \in \Omega_i \Leftrightarrow \max_i Y_{\tilde{A}_t}(\omega, \Omega_i), \qquad (5)$$

где $i = \overline{1,m}$, $t = \overline{1,t_{\max}}$, $Y_{\tilde{A}_t}(\omega, \Omega_i)$ – описаний объектов, определяемая по формуле

$$Y_{\tilde{A}_t}(\omega, \Omega_i) = \sum_{t=1}^{t_{\max}} \tilde{A}_t(\omega, \Omega_i), \qquad (6)$$

либо эта оценка будет превышать оценки всех остальных классов не меньше чем на определенную пороговую величину λ и т.д.

Однако этот метод при определенном достоинстве, таком как простота исполнения, все же обладает недостатком, это значительное число машинных операций при большой мощности словаря признаков, что снова не приемлемо в условиях ограничения временного и вычислительного ресурса.

Таким образом, рассмотренные методы не способны эффективно применяться в условиях быстроменяющейся информационной обстановки.

Решим эту проблему следующим образом. Пусть, как в предыдущих случаях, существует словарь признаков (таблица), в котором на языке этих признаков описаны объекты, составляющие классы распознавания. Причем условием распознавания является то, что ошибка измерения меры близости неопознанного объекта к классу не должна превышать величины $R(\Omega_{i-1}, \Omega_i)$ – меру близости между классами, определяемую по формуле

$$R(\Omega_{i-1}, \Omega_i) = \sqrt{\frac{1}{r_{i-1} r_i} \sum_{v=1}^{r_{i-1}} \sum_{t=1}^{r_i} d^2(\omega_{i-1v}, \omega_{it})}, \qquad (7)$$

где $i = \overline{1,m}$, $d^2(\omega_{m-1v}, \omega_{mt})$ – расстояния между объектами соседних классов,

r – количество объектов в классе.

Выделим дополнительно в словаре признаков границы класса распознавания, состоящие из объектов класса, которые описываются экстремальными (максимальными или минимальными) значениями по одному и более признакам.

Пример описания границ класса в двухмерном информационном пространстве показан на (рис. 2).

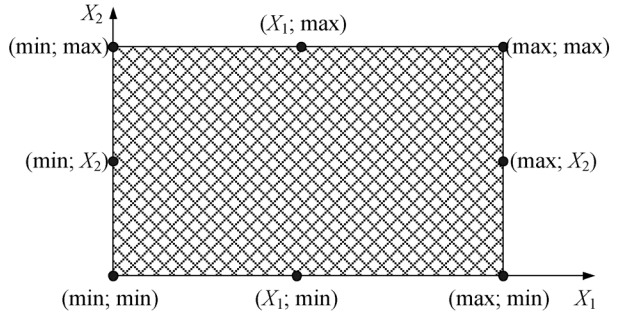

Рис. 2. Пример описания границ класса распознавания в двухмерном информационном пространстве

При появлении неопознанного вида воздействий в информационном N-мерном пространстве I, меру близости L_i, где $i=\overline{1,m}$, определяем путем вычисления расстояния до границы класса (рис. 3).

Для этого по формуле (1) вычисляем расстояние $d^2_{i\tilde{a}pl}$, где $i=\overline{1,m}$ от объекта распознавания ω до граничного объекта $\omega_{i\tilde{a}p\min}$, где $i=\overline{1,m}$ с координатами, имеющими минимальные значения для данного класса распознавания Ω_i, где $i=\overline{1,m}$. Далее вычисляем расстояние $d^2_{i\tilde{a}p2}$, где $i=\overline{1,m}$ от объекта распознавания ω до граничного объекта $\omega_{i\tilde{a}p\max}$, где $i=\overline{1,m}$ с координатами, имеющими максимальные значения для данного класса распознавания Ω_i, где $i=\overline{1,m}$.

Сравниваем полученные расстояния $d^2_{i\tilde{a}pl}$ и $d^2_{i\tilde{a}p2}$, где $i=\overline{1,m}$. В случае, если $d^2_{i\tilde{a}pl}<d^2_{i\tilde{a}p2}$, то вычисляем $d^2_{i\tilde{a}\delta2+k_z}$ расстояния, где $i=\overline{1,m}$, $k_z=\overline{1,C_N^{s_q}}$, значение $C_N^{s_q}$ находим по формуле

$$C_N^{s_q}=\frac{N!}{(N-s_q)!s_q!}. \qquad (8)$$

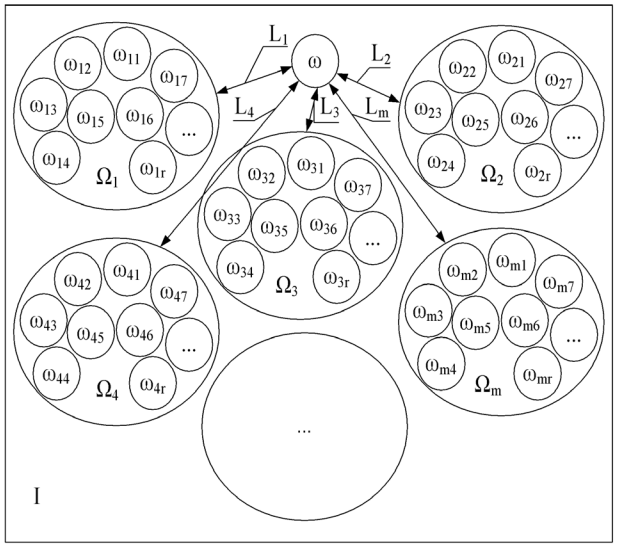

Рис. 3. Определение меры близости до границы класса

Значения величин $d^2_{i\tilde{a}p2+k_z}$ – это расстояния до граничных объектов $\omega_{i\tilde{a}p*}$, где $i=\overline{1,m}$ с координатами, имеющими $s_q=1$ значений максимальными, остальные минимальными для данного класса распознавания Ω_i, где $i=\overline{1,m}$. Сравниваем полученные расстояния и определяем минимальное из них.

После этого вычисляем расстояния $d^2_{i\tilde{a}p2+k}$, где $i=\overline{1,m}$, k находим по формуле

$$k=\left(\sum_{g=1}^{y}C_N^g\right)+k_{z+y}, \qquad (9)$$

где $k_{z+y}=\overline{1,C_N^{s_{q+y}}}$, $C_N^{s_{q+y}}$ находим согласно выражения (8), в котором $s_{q+y}=s_q+y$, $y=\overline{1,(N-2)}$.

Полученные величины – это расстояния до граничных объектов $\omega_{i\tilde{a}p*}$, где $i=\overline{1,m}$ с координатами, имеющими $s_{q+y}=s_q+y$, где $y=\overline{1,(N-2)}$ значений максимальными, остальные минимальными для данного класса распознавания Ω_i, где $i=\overline{1,m}$. Сравниваем полученные расстояния и определяем минимальное из них.

В случае, если $d^2_{i\tilde{a}pl}>d^2_{i\tilde{a}p2}$, где $i=\overline{1,m}$, то вычисляем расстояния $d^2_{i\tilde{a}p2+k_z}$, где $i=\overline{1,m}$,

$k_z = \overline{1, C_N^{s_q}}$, значение $C_N^{s_q}$ определяем по формуле (8).

Полученные значения величин являются расстоянием до граничных объектов $\omega_{i\text{гр}*}$, где $i = \overline{1,m}$ с координатами, имеющими $s_q = 1$ значений минимальными, остальные максимальными для данного класса распознавания Ω_i, где $i = \overline{1,m}$. Сравниваем полученные расстояния и определяем минимальное из них.

Далее вычисляем расстояния $d^2_{i\text{гр}2+k}$, где $i = \overline{1,m}$, k находим по формуле (9).

Полученные величины – это расстояния до граничных объектов $\omega_{i\text{гр}*}$, где $i = \overline{1,m}$ с координатами, имеющими $s_{q+y} = s_q + y$, где $y = \overline{1,(N-2)}$ минимальные, остальные максимальные для данного класса распознавания Ω_i, где $i = \overline{1,m}$. Сравниваем полученные расстояния и определяем минимальное из них.

В результате полученных вычислений получаем N расстояний от объекта распознавания ω до граничных объектов класса $\omega_{i\text{гр}*}$, где $i = \overline{1,m}$ все эти объекты имеют одну общую координату x_p, где $p = \overline{1,N}$ и отличаются друг от друга также по одной координате x_{p*}, где $p = \overline{1,N}$. Запоминаем значение общей координаты.

Далее рассматриваем пару граничных объектов, отличающихся друг от друга только по одной координате x_{p*}, где $p = \overline{1,N}$, находим значение этой координаты для ближайшего объекта, принадлежащего классу распознавания по формуле

$$x_{p*} = \frac{d^2_{i\text{гр}}(..,x_{p\min},..) + a^2 - d^2_{i\text{гр}}(..,x_{p\max},..)}{2 \cdot a^2} + x_{p\min}, \quad (10)$$

где $a = (x_{p\max} - x_{p\min})$ – расстояние между объектами, принадлежащими классу распознавания и отличающимися только по одной координате, $d^2_{i\text{гр}}(..,x_{p\min},..)$ и $d^2_{i\text{гр}}(..,x_{p\max},..)$ – расстояние от объекта распознавания ω до граничных объектов класса $\omega_{i\text{гр}*}$, где $i = \overline{1,m}$ отличающихся только по одной координате и имеющих минимальное $x_{p\min}$ и максимальное $x_{p\max}$, где $p = \overline{1,N}$ значение этой координаты соответственно.

Данную процедуру повторяем для всех пар граничных объектов, отличающихся друг от друга только по одной координате.

Таким образом, мы получим координаты граничного объекта $\omega_{i\text{гр}*}$, где $i = \overline{1,m}$ находящегося на наименьшем удалении от объекта распознавания ω. Вычислим расстояние $d^2_{i\text{гр}*}$, где $i = \overline{1,m}$ от объекта распознавания ω до найденного граничного объекта $\omega_{i\text{гр}*}$, где ω, найденное значение $d^2_{i\text{гр}*}$, где $i = \overline{1,m}$ является искомой величиной L_i, где $i = \overline{1,m}$.

Далее на основании близости $L = \min\{L_i\}$, где $i = \overline{1,m}$, к какому-либо классу принимается решение о принадлежности рассматриваемого объекта ω, негативного воздействия, к этому классу Ω_i, где $i = \overline{1,m}$.

Решающее правило выглядит так же, как в методе, основанном на использовании алгоритма минимума расстояния (формула (3)).

Однако преимущества, в отличие от вышеизложенных методов, очевидны. Во-первых, уменьшается количество машинных операций, для определения принадлежности неизвестного объекта к определенному классу, так как предложенный способ не зависит от количества объектов составляющих класс. На рис. 4 зависимости 1 и 2 соответствуют применению классического метода распознавания при Ω_i=6, ω_j=300 и ω_j=1000 соответственно, где $i = \overline{1,m}$, $j = \overline{1,r}$. Зависимость 3 на рис. 4 соответствует применению предлагаемого способа распознавания при Ω_i=6 и ω_j=1000, где $i = \overline{1,m}$, $j = \overline{1,r}$.

Во-вторых, при неизменной мощности словаря признаков, время на распознавания значительно сокращается, так как, в отличие от остальных методов, в предложенном способе рассматриваются лишь объекты, находящиеся на границе класса и совершенно не учитываются объекты, принадлежащие классу, но находящиеся глубже по параметрам признакового пространства.

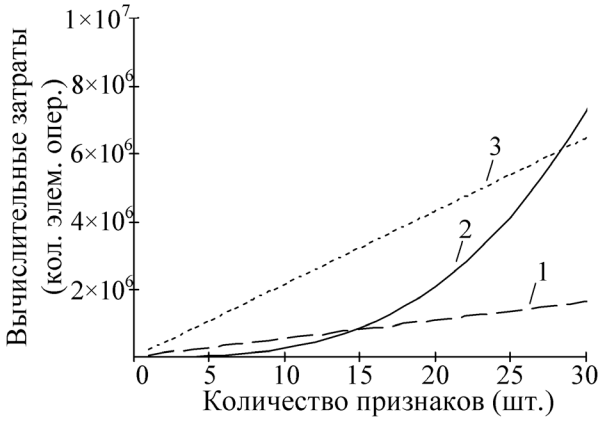

Рис. 4. График зависимости вычислительных затрат от количества объектов и признаков распознавания

Таким образом, анализируя все достоинства предложенного способа, можем сказать, что с его применением появляется возможность использовать высвободившийся ресурс (временной и вычислительный) для увеличения количества анализируемых параметров негативных воздействий, а это, в свою очередь, увеличит ситуационную адаптивность СЗИ в условиях интенсивного информационного противоборства.

Литература

1. Основы информационной безопасности: учебник для высших учебных заведений МВД России / под ред. В.А. Минаева и С.В. Скрыля. - Воронеж: Воронежский институт МВД России, 2001. - 464 с.

2. Методы и средства автоматизированного управления подсистемой контроля целостности в системах защиты информации: монография / Е.А. Рогозин, А.С. Дубровин, В.И. Сумин и др. // Монография. – Воронеж: Воронеж. гос. техн. ун-т, 2003. – 165 с.

3. Миленький А.В. Классификация сигналов в условиях неопределенности. Статистические методы самообучения в распознавании образов /А.В. Миленький - М.: Советское радио, 1975. – 328 с.

4. Фукунага К. Введение в статистическую теорию распознавания образов/К.Фукунага – М.: Наука, 1979. – 368 с.

.

Воронежский институт МВД России
The Voronezh institute of the Ministry of Internal Affairs of Russia

METHOD FOR IDENTIFYING NEGATIVE IMPACT ON THE INFORMATION SYSTEMS INFORMATION SECURITY

S.V. Skryl', S.V. Belokurov, A.I. Boronenkov, P.E. Krasnov

Existing methods of identification of negative consequences are considered, and the assessment of their functioning in the conditions of negative impacts in the presence of rigid restrictions on a time and computing resource of system of protection of information is made. The new way of identification of negative impacts is offered and its assessment is made

Key words: recognition, class, sign, object, method, distance, proximity measure, border, information space