Методы интеллектуального анализа событий информационной безопасности в информационно-телекоммуникационных сетях

Зубков Е.В.

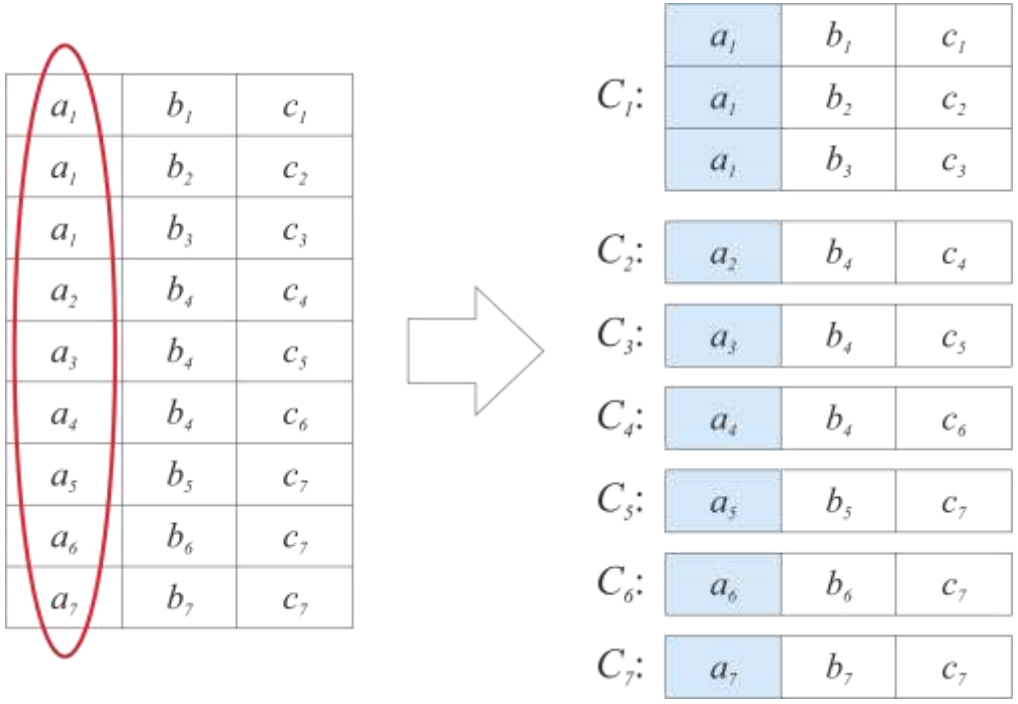
Новосибирск
2016

Проблематика

При работе специалиста с большими объемами статистической информации часть данных неизбежно выпадает из поля зрения.

Подходы к работе с данными

Группировка данных на основе
всех значений одного (нескольких) признаков



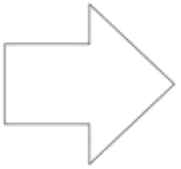
	a_1	b_1	c_1
	a_1	b_2	c_2
	a_1	b_3	c_3
	a_2	b_4	c_4
	a_3	b_4	c_5
	a_4	b_4	c_6
	a_5	b_5	c_7
	a_6	b_6	c_7
	a_7	b_7	c_7

C_1:	a_1	b_1	c_1
	a_1	b_2	c_2
	a_1	b_3	c_3
C_2:	a_2	b_4	c_4
C_3:	a_3	b_4	c_5
C_4:	a_4	b_4	c_6
C_5:	a_5	b_5	c_7
C_6:	a_6	b_6	c_7
C_7:	a_7	b_7	c_7

Подходы к работе с данными

Группировка данных на основе
отдельных значений всех признаков

a_1	b_1	c_1
a_1	b_2	c_2
a_1	b_3	c_3
a_2	b_4	c_4
a_3	b_4	c_5
a_4	b_4	c_6
a_5	b_5	c_7
a_6	b_6	c_7
a_7	b_7	c_7



C_1:

a_1	b_1	c_1
a_1	b_2	c_2
a_1	b_3	c_3

C_2:

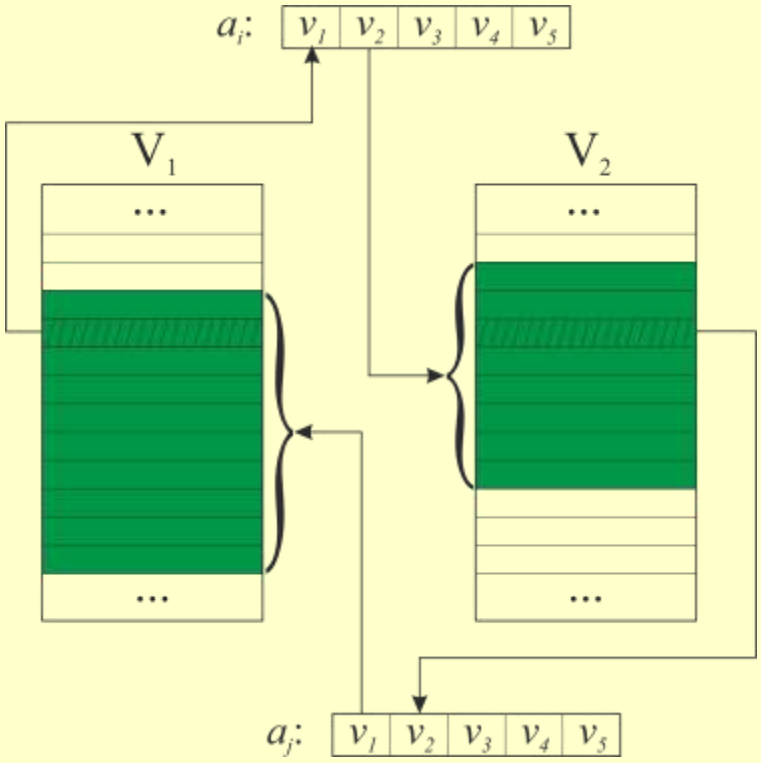
a_2	b_4	c_4
a_3	b_4	c_5
a_4	b_4	c_6

C_3:

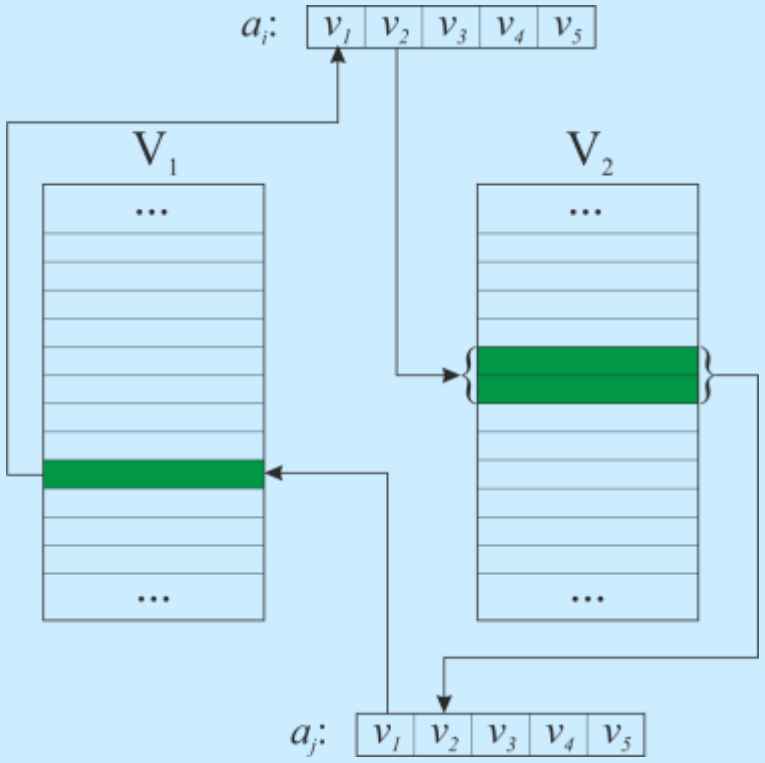
a_5	b_5	c_7
a_6	b_6	c_7
a_7	b_7	c_7

Информативность признака

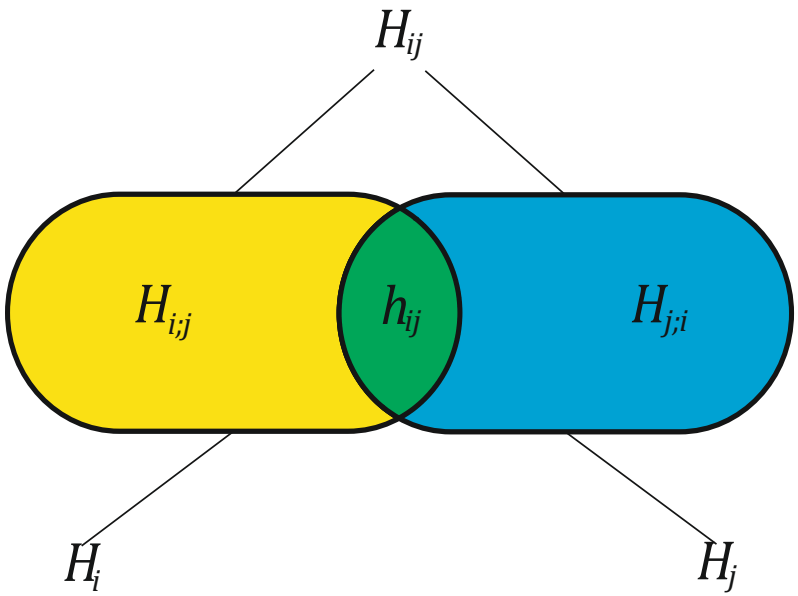
Пример 1



Пример 2



Энтропийный анализ данных



H_i - энтропия i-го признака
H_j - энтропия j-го признака
H_{ij} - энтропия сочетаний i-го и j-го признаков

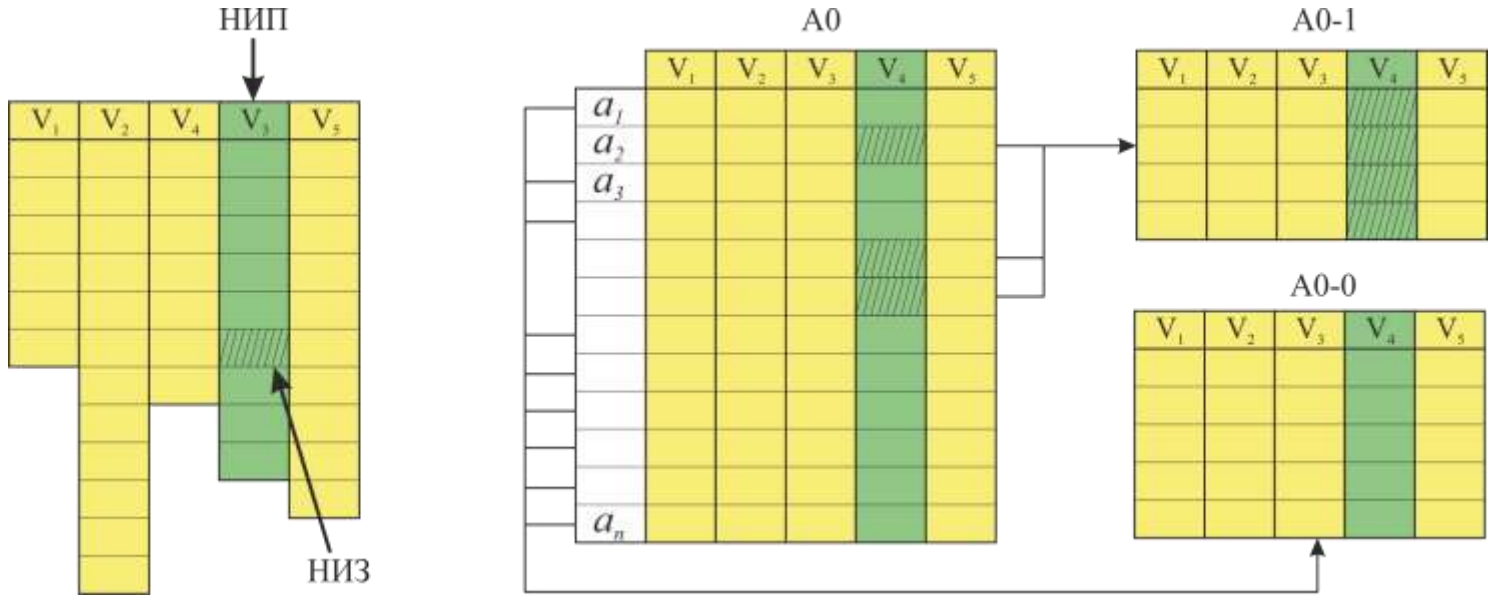
$$h_{ij} = H_i - H_{j;i} = H_i + H_j - H_{ij} = H_i - H_{i;j}$$

$\frac{h_{ij}}{H_{ij}}$ - коэффициент взаимного влияния признаков v_i и v_j

$I_i = \frac{\sum_{j \neq i} h_{ij}}{\sum_{j \neq i} H_{ij}}$ - информативность признака v_i

Энтропийный анализ данных

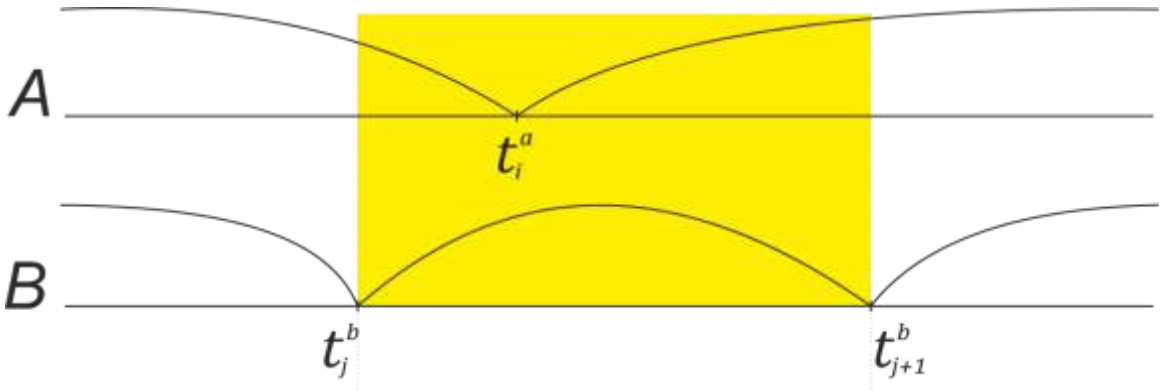
- Вычисление наиболее информативного признака (НИП)
- Вычисление наиболее информативного значения (НИЗ)
- Выбор из исходного множества элементов у которых значение НИП соответствует НИЗ



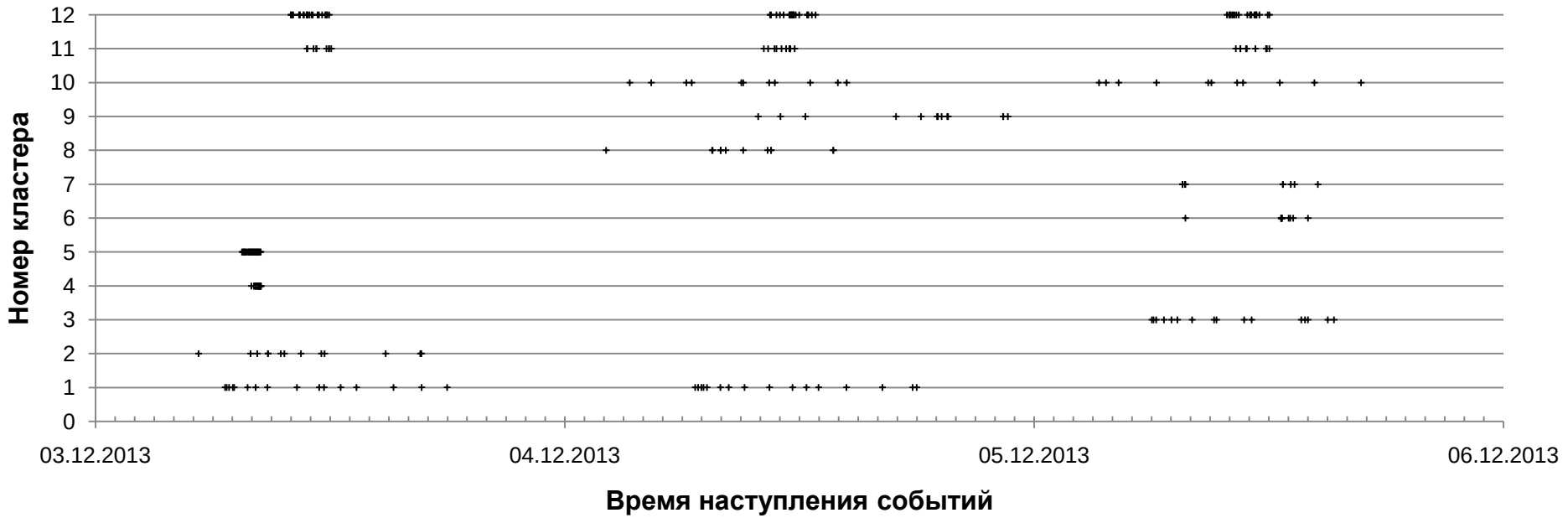
Кластер как последовательность событий

Распределение событий кластера во времени

Соотношение событий двух кластеров



Вычислительный эксперимент



i	*j*	*I*
4	5	0,883
6	7	0,906
11	12	0,840
3	6	0,661
3	7	0,651

Оценка степени угроз

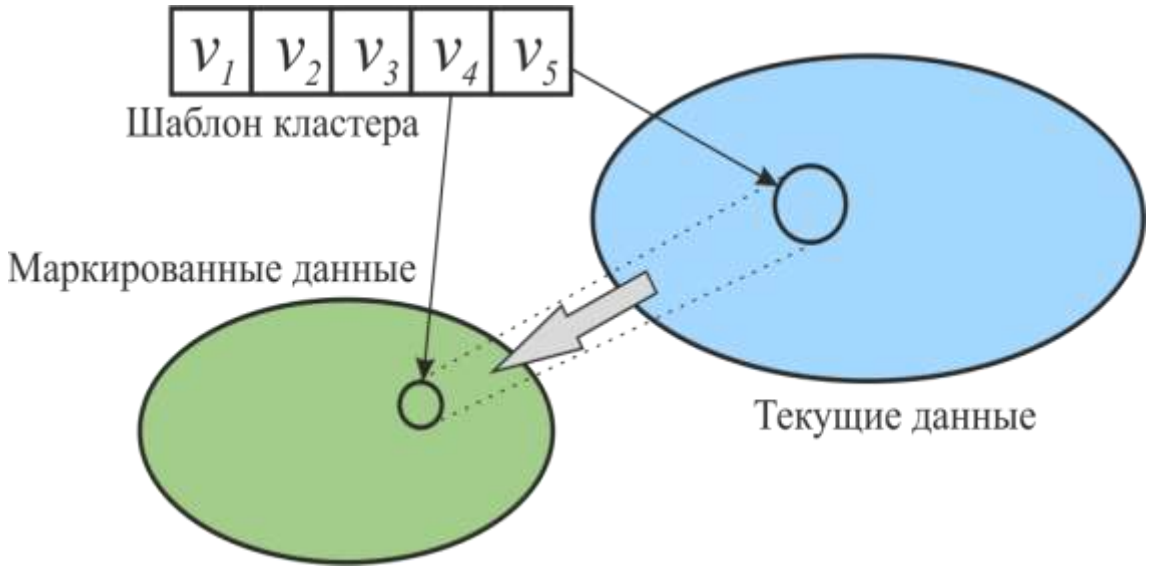
Степень изученности кластера:

$$\begin{cases} R = 1, \text{при } N_m \neq 0, N_{\bar{m}} = 0; \\ R = \dfrac{N_m \cdot U_{\bar{m}}}{N_s \cdot (U_{\bar{m}} + U_m)}, \text{при } N_m \neq 0, N_{\bar{m}} \neq 0\ ; \\ R = 0, \text{при } N_m = 0, N_{\bar{m}} \neq 0. \end{cases}$$

N_m – количество маркированных элементов,
$N_{\bar{m}}$ – количество немаркированных элементов,
U_m – однородность множества маркированных элементов,
$U_{\bar{m}}$ – однородность множества немаркированных элементов.

Оценка уровня совокупной угрозы:

$$R_s = 1 - \prod_{i=1}^{N}(1 - R_i)$$



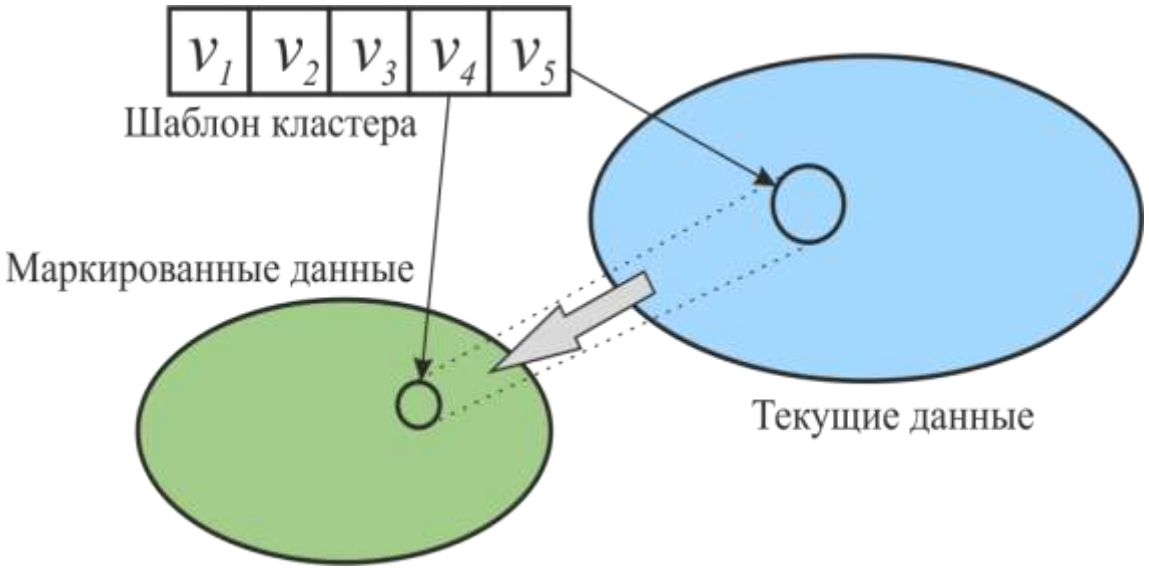
Оценка степени угроз

Показатель непротиворечивости экспертной оценки:

$$C = (1 + P_{m+} \cdot \log P_{m+} + P_{m-} \cdot \log P_{m-}) \cdot sign$$

$$sign = \begin{cases} 1, & n_{m+} \geq n_{m-} \\ -1, & n_{m+} < n_{m-} \end{cases}.$$

P_{m+} – вероятность положительного заключения,
P_{m-} – вероятность отрицательного заключения,
n_{m+} – количество положительных заключений,
n_{m-} – количество отрицательных заключений.



Вывод

Предлагаемый подход позволяет:

- Существенно сократить количество сущностей, требующих экспертного анализа.
- Выявлять динамическую зависимость между кластерами.
- Получить новые признаки статистического характера, которые могут быть использованы для дальнейшего анализа, в том числе и в автоматическом режиме.
- Получить вероятностную оценку соответствия текущей сетевой активности действительной угрозе.

Спасибо за внимание !