

Data Science для решения задач информационной безопасности

## Лекция 1. Экспертные системы, полнота, точность. Задачи ИБ, решаемые ЭС.

---

Павел Владимирович Слипенчук

17 сентября 2019

Москва, МГТУ им.Бауманка,  
каф.ИУ-8, КИБ

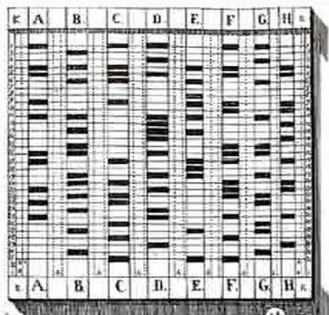
1. История. Заметки
2. Экспертная система. Априорная и апостериорная вероятности. Полнота, точность экспертной системы
3. Задачи ИБ, решаемые экспертными системами

## История. Заметки

---

# Гомеоскоп Корсакова

Семён Павлович Корсаков (1787-1853) – изобретатель первых экспертных систем (механических): гомеоскопов, применявшихся в медицине.



Домашнее задание:

<https://ru.wikipedia.org/wiki/Гомеоскоп>

- «Паскалина» – суммирующая машина Блеза Паскаля (1642).
- Разностная машина Иоганна Мюллера (1788, не построена)
- Разностная машина Чарльза Бебиджа (малая 1822, построена; большая 1849, построена в 2000)
- Гомеоскоп С.Н.Корсакова (1832)

- Советский физиолог Пётр Кузьмич Анохин формулирует понятие «обратной связи» (1935)
- Взлом шифровальной машины «Энигма» (Мариан Реевски). Проект «Бомба» (Ежи Рожицки, Генрих Зыгальски). Передача чертежей и спецов в Кабацком лесу под Варшавой английской разведке (1936). Проект «Колосс» (под руководством Алана Тьюринга, 1936-1941)
- Фашистский антикварный проект (первый гипертекст, 1937)
- Первые ЭС для военных целей (зенитные орудия, ТАУ, 1939).
- Системы автоматического регулирования (САР), Владимира Викторовича Солодовникова (1939)

- Машина Конрада Цузе (1941) и высокоуровневый язык программирования: Планкалкюль (1948)
- ENIAC Джона Мокли и Джона Преспера Эккерта (1943, 1946 публично представлен)
- «Кибернетика» Норберта Винера (1948)

- ОГАС<sup>1</sup> и ЕГСВЦ<sup>2</sup> Анатолия Ивановича Китова и Виктора Михайловича Глушкова. (1958-1964).
- МИР<sup>3</sup> Глушкова – первый ПК.

(!) В 1964 году выходит статья в The Washington Post «Перфокарта управляет Кремлём», после которой Политбюро ЦК КПСС принимает решение о сворачивании проекта ОГАС.

## Д3. Проект «Киберсин» в Чили.

---

<sup>1</sup>Общегосударственная Автоматизированная Система

<sup>2</sup>Единая Государственная Сеть Вычислительных Центров

<sup>3</sup>Машина для инженерных расчётов



- Метод опорных векторов (Support vector machine) Алексея Яковлевича Червоненкиса и Владимира Наумовича Вапника (1963).
- Случайный лес (Random Forest) Юрия Леонидовича Павлова (1977)
- Bootstrap Бредли Эйфон (1977)
- Байесовские сети Джуды Перла Judea Pearl (1988)
- Bagging (Bootstrap Aggregating) Лео Бреймана (1994)
- Boosting Роберта Шапира (1990)

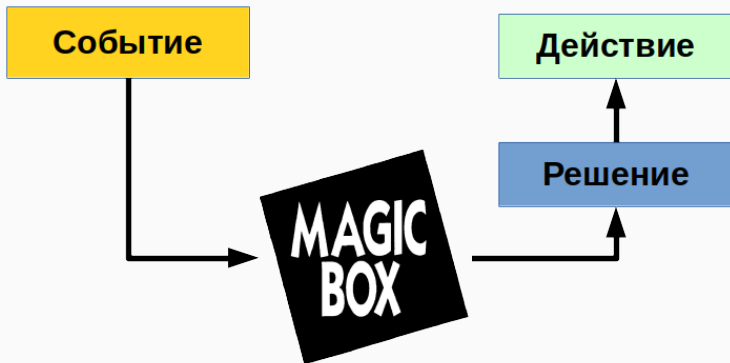
- “Закон Мура” и появление доступных персональных компьютеров открыло новую эпоху в ML и DM (1990 – н.в.)
- первая<sup>4</sup> NoSQL СУБД “Strozzi NoSQL” Карло Стрози (1998)

---

<sup>4</sup>после появления реляционной алгебры

Экспертная система. Априорная и апостериорная вероятности.  
Полнота, точность экспертной системы.

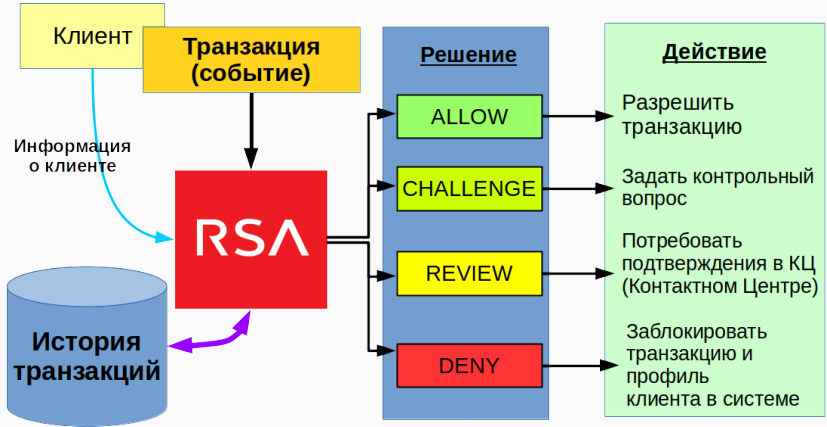
---



Процесс:

1. Событие
2. Экспертная система
3. Решение (экспертной системы)
4. Действие (основанное решением экспертной системы)

# ФМ система RSA



Есть открытое описание бизнес-процессов в Хакере:

<https://xakep.ru/2017/05/22/bank-antifraud-uncovered/>

### **Замечание.**

Для улучшения качества работы экспертной системы, понимания границ ее применения, необходимо знать о её внутреннем устройстве.

Но для оценки её качества на тестовых данных, ЭС можно воспринимать как чёрный ящик. Таким образом, чтобы принять работу, не требуется быть специалистом в области DS/ML.

**Априорная вероятность** – вероятность взятая из каких-либо умозаключений или правил. Примеры:

1. Вероятность выпадения орла 0.5, потому что он ничем не лучше и не хуже решки и монетка не может упасть ребром (это пренебрежимо мало)
2. Мы отправили *событие* на вход ЭС и получили на выходе решение: "вероятность мошенничества равна 0.7 для данного события".

**Апостериорная вероятность** – статистическая вероятность<sup>5</sup>, посчитанная на каких-либо конкретных данных.

Примеры:

1. Мы 100 раз подбросили монету и 47 раз выпал орёл.  
Следовательно вероятность выпадения орла 0.47
2. Мы взяли экспертную систему и посчитали, что "вероятность мошенничества 0.7" выпало на 10 мошеннических и 100 легитимных операций за определённое время. Значит *точность* системы для данного *решения* на данном промежутке времени равна 0.1

---

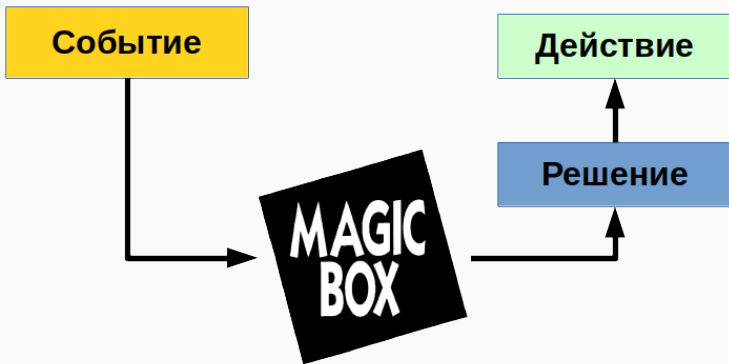
<sup>5</sup>Вообще то говоря, апостериорная вероятность имеет более глубокий и широкий смысл. Но в рамках нашего курса апостериорная вероятность – это просто статистическая доля того или иного события.



Современные ЭС состоят из двух блоков:

1. ***Score Engine*** – программный модуль, выдающий значение скоринга (как правило от 0 до 1000, от 0.0 до 1.0 или от -1.0 до 1.0). Чем выше это значение, тем больше *априорная вероятность* что произошёл какой-либо инцидент. Score Engine выдает либо одно значение (RSA) либо несколько (Secure Bank, Group-IB).
2. ***Rules Engine*** – программный модуль, использующий данные Score Engine и другие данные и выдающий *решение* экспертной системы.

Иногда ЭС называют только Score Engine. Таким образом получаем большое разнообразие решений вида:  
"априорная вероятность мошенничества равна X".



Как понять, как хорошо работает экспертная система?

Ошибка I рода

$\alpha$  ошибка

false positive (fp)

Ложная сработка

Ошибка II рода

$\beta$  ошибка

false negative (fn)

Пропуск цели

# Четыре События

На примере задачи фрод-мониторинга в банковской сфере.

1.  $F_r$  (fraud real) – событие, означающее что банковская операция в действительности является мошеннической.
2.  $F_s$  (fraud system) – событие, когда некая система фрод-мониторинга оценила данную операцию как мошеннической.
3.  $L_r$  (legitim real) – событие, означающее что банковская операция в действительности является легитимной (не мошеннической).
4.  $L_s$  (legitim system) – событие когда некая система фрод-мониторинга оценила данную операцию как легитимную (не мошеннической).

События  $F_s$  и  $L_s$  несовместны, а так же  $F_r$  и  $L_r$  несовместны.

Таким образом можно рассматривать следующие события:  
 $P(F_r \cap F_s), P(F_r \cap L_s), P(L_r \cap F_s), P(L_r \cap L_s)$ . Для удобства будем обозначать их просто:  $P(F_r F_s), P(F_r L_s), P(L_r F_s), P(L_r L_s)$ .

Так же имеет смысл рассматривать условные вероятности:  
 $P(F_r|F_s), P(F_s|F_r), \dots$ .

## Ложная сработка & пропуск цели

Ошибка I рода

$\alpha$  ошибка

false positive (fp)

Ложная сработка

$$P(L_r|F_s)$$

Ошибка II рода

$\beta$  ошибка

false negative (fn)

Пропуск цели

$$P(F_r|L_s)$$

### Замечание.

Иногда false positive называют событие  $P(L_r F_s)$ , а не  $P(L_r|F_s)$ . Аналогично false negative называют  $P(F_r L_s)$ , а не  $P(F_r|L_s)$ . Это очень важно! Не запутайтесь!

## Полнота & точность

Основными показателями качества системы являются *полнота* ( $\Pi$ , *recall*) и *точность* ( $T$ , *precision*).

Их определения:

$$\Pi \stackrel{\text{def}}{=} P(F_s|F_r) \quad (1)$$

$$T \stackrel{\text{def}}{=} P(F_r|F_s) \quad (2)$$

Полноту и точность можно вычислить через формулы:

$$\Pi = \frac{P(F_s F_r)}{P(F_s F_r) + P(L_s F_r)} \quad (3)$$

$$T = \frac{P(F_s F_r)}{P(F_s F_r) + P(F_s L_r)} \quad (4)$$

**ДЗ: докажите эти формулы**

## Использование формул (3) и (4)

Формулы (3) и (4) полезны для определения полноты и точности через статистические данные.

Предположим, что в каком-либо банке за определенный промежуток времени произошло **562** мошеннических транзакций. Выберем их все и случайным способом ещё **100 000** легитимных транзакций. Предположим что всего за этот промежуток произошло 67 234 134 234 операций.

Тогда наша выборка легитимных операций это  $1/q$  от общей доли, где:

$$q = \frac{67234134234}{100000} \simeq 672 \cdot 10^3$$

Мы прогнали все эти транзакции через ЭС и обнаружили количество пар событий:  $C(F_S F_r) = 386$ ,  $C(F_S L_r) = 18$ ,  $C(L_S F_r) = 176$ ,  $C(L_S L_r) = 999982$ .

... ..



## Использование формул (3) и (4)

... ..

Таким образом через формулу (4) можно найти точность:

$$\begin{aligned} T &= \frac{P(F_S F_R)}{P(F_S F_R) + P(F_S L_R)} = \frac{C(F_S F_R)}{C(F_S F_R) + q \cdot C(F_S L_R)} = \\ &\simeq \frac{386}{386 + 18 \cdot 672 \cdot 10^3} \simeq \frac{386}{18 \cdot 672 \cdot 10^3} \simeq 3 \cdot 10^{-5} \end{aligned}$$

### Замечание

Очень часто при подсчёте точности забывают про коэффициент  $q$  и совершают ошибку, измеряя точность на конкретных данных. На практике физически невозможно выбрать все легитимные транзакции, поэтому берут их подмножество.

**ДЗ: найдите полноту**

Ошибки I и II рода можно выразить через полноту и точность.

$$O_2 = 1 - \Pi \quad (5)$$

$$O_1 = \frac{P(F_r)}{P(L_r)} \cdot \Pi \cdot \left( \frac{1}{T} - 1 \right) \quad (6)$$

ДЗ. Докажите формулы (5) и (6)

ДЗ. Выразите полноту и точность, через  $O_1$  и  $O_2$

Пропуск цели

vs

Ложная сработка

**Что важнее?**

ЭС диагностики раковой опухоли:

- **Ложная сработка** – решение ЭС в том, что пациент болен; но на самом деле пациент здоров.
- **Пропуск цели** – система говорит больному раком пациенту, что он здоров.

ЭС обнаружения превышения скорости автомобиля:

- **Ложная сработка** – штраф будет выписан добропорядочному водителю.
- **Пропуск цели** – лихач не будет наказан за превышение скорости и ему не выпишут штраф

Фрод мониторинг

- **Ложная сработка** – заблокировать легитимную транзакцию.
- **Пропуск цели** – позволить хакерам украсть деньги клиента

На практике: определяют допустимое количество нагрузки на контактный центр и забивают эту нагрузку «под завязку».

Отсюда выводы: мелкие хищения менее интересны чем крупные.

# Булевый и вероятностный отклик

**Булевый отклик** – либо 0, либо 1. Либо -1, либо 1. (виновен/не виновен).

**Вероятностный отклик** – классификатор выдает вероятностное *априорное* решение  $p \in [0, 1]$ : 0 – нет, 1 – да, 0.5 – неопределенно. (или  $p \in [-1, 1]$ : -1 – нет)

Любое вероятностное решение можно свести к булевому. Для этого определяется некое значение, называемое **отсечкой** (cutoff):

$$p \geq \text{cutoff} \implies \text{return } 1 \quad (7)$$

$$p < \text{cutoff} \implies \text{return } 0$$

## Замечание

Значение  $p$  – это *априорные вероятности* ЭС. Они могут не иметь никакого отношения к реальности

## Влияние отсечки на полноту и точность

Было:

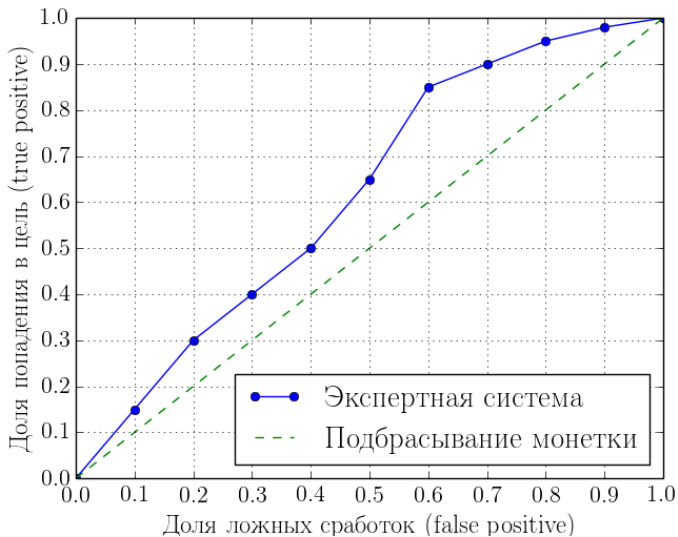
$p \geq 0.5$  – виновен;  $< 0.5$  – не виновен

Поменяли отсечку. Стало:

$p \geq 0.8$  – виновен;  $< 0.8$  – не виновен

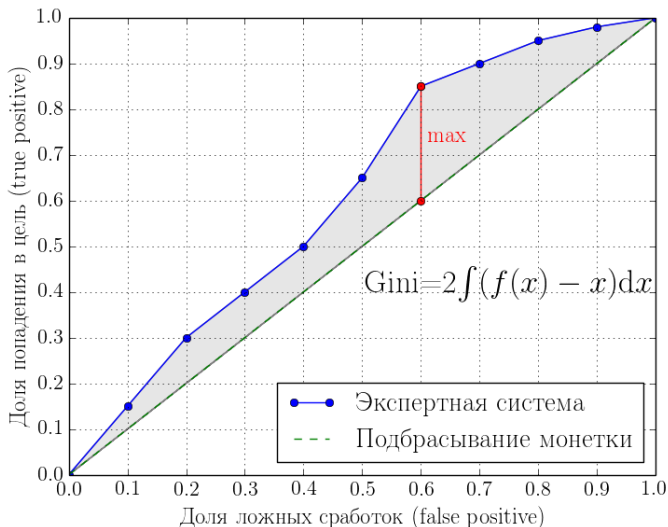
Что стало с полнотой и точностью? Что стало с ложной сработкой и пропуском цели? Что не увеличилось, а что не уменьшилось?

## Receiver Operating Characteristic

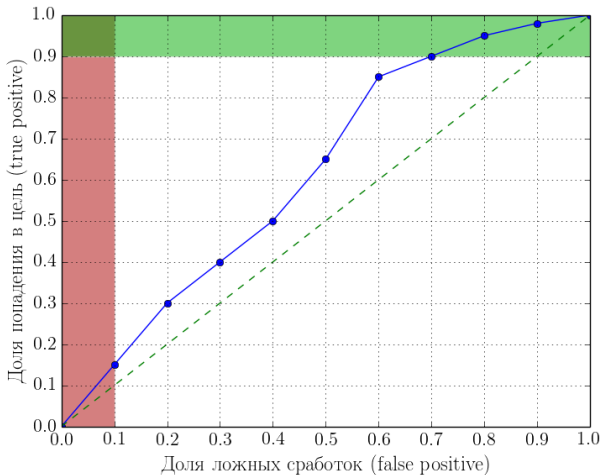




# Коэффициент Джини & Максимальное отклонение («Расстояние Робина Гуда»)

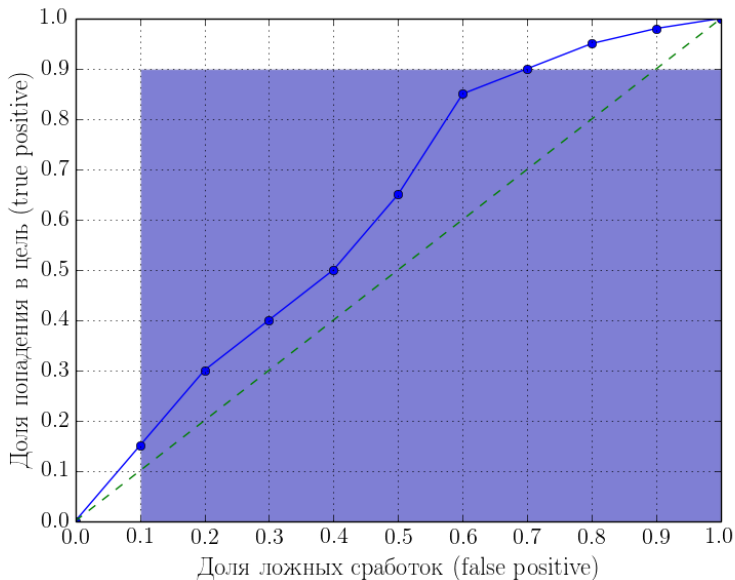


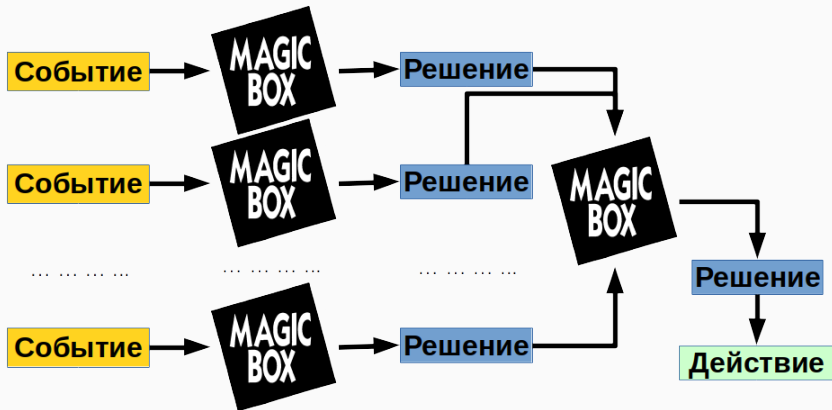
# Используемые области



Какие области используются для детектирования рака и для обнаружения превышения скорости автомобилей?

А эти области никогда не используются???





# Задачи ИБ, решаемые экспертными системами

---

- Банковский фрод
- Криминалистика
- Защита бренда, антипиратство, антиконтрафакт
- NGAV
- Дактилоскопия
- Анализ трафика
- Информационные войны
- Pay per click фрод
- Call фрод
- Анализ даркнета
- Противодействие спаму
- ...

- персональные данные
- вкусовые предпочтения, речь, увлечения (социальные сети)
- поведение
- mouse track analysys
- анализ данных телефона
- keystroke dynamics
- предпочтения
- фотография, голос, видео
- файлы
- данные трафика
- ...

## Вопросы для самопроверки

---



- Вася посчитал, что в ЭС, обнаруживающей мошенничество в банковской сфере, доля ложных сработок всего 0.001% от общего числа транзакций. Вася внедрил систему. Васю уволили. Почему?
- Петя посчитал, что в ЭС детектирования наличия диабета у пациентов, которые подозревают у себя его наличие, доля ложных сработок всего 0.1% от общего числа операций. Петя внедрил систему. Петю повысили. Почему?
- Коля посчитал, что в ЭС детектирования наличия диабета у школьников, сдавшие «добровольно-принудительные» анализы, всего 0.1% от общего числа школьников. Коля внедрил систему. Колю уволили. Почему?
- Джон посчитал, что в ЭС детектирования наличия диабета у школьников, сдавшие принудительные анализы, всего 0.1% ложных сработок от общего числа школьников. Но в отличие от Коли, он работает в США. Джон внедрил систему. Джона повысили, в отличие от Коли. Почему?

1. Используя формулу Байеса, докажите формулы (3) и (4) из слайда №20
2. Найдите полноту в примере на слайде №21.
3. Докажите формулы (5) и (6): вычисление ошибок первого и второго рода через полноту и точность
4. Вычислите полноту и точность через ошибки первого и второго рода
5. Почему специалисты Data Science используют полноту и точность, но редко пользуются ошибками первого и второго рода?
6. классификатор выдает решение  $p \in [0, 1]$ , однако для решения определённой задачи хотелось бы чтобы классификатор выдавал решение  $p \in [-1, 1]$ .  
Предложите простой фильтр.