

Презентация к Госэкзамену

Луничкин Е.В.

Московский физико-технический институт (государственный университет)

Москва – 2019

Содержание

- 1 История создания КИС в США, СССР, России
- 2 ВФР
- 3 Теория тезауруса
- 4 Диплом

История создания КИС в США, СССР, России. Развитие КИС: этапы, терминология. Структура базовых КИС. Современное состояние базовых КИС. Тенденции развития

История развития КИС (1/4)

- **1904-49 гг., теория** – принципы организации производства, заложенные Ф. У. Тейлором (F. W. Taylor, 1856-1915), пока без аппаратно-программной реализации (в силу «неизобретённости» и неразвитости вычислительной техники).
- **1950-64 гг., IC/IM (MRP 0)** – оптимизация складских запасов (Inventory Control/Management, IC/IM), расчёт и планирование потребностей в материалах (Material Requirements Planning, условно MRP 0) по Дж. Орлики (J. Orlicky) и О. Уайту (O. Wight).
- **1965-74 гг., MRP** – планирование потребностей в материалах (Material Requirements Planning, MRP или MRP 1), в том числе по замкнутому циклу (Closed Loop MRP), включающее составление производственной программы и ее контроль на цеховом уровне по Дж. Дж. Миллеру (J. G. Miller) и Л. Дж. Спраг (L. G. Sprague).

История развития КИС (2/4)

- **1975-80 гг., MRP 2** – планирование производственных ресурсов (Manufacturing Resources Planning, MRP 2) на основе данных, полученных от поставщиков и потребителей, включая прогнозирование, планирование (в том числе загрузки производственных мощностей) и контроль за производством.
- **1981-85 гг., CALS** – добавление к MRP 2 идеологии «точно в срок» (Just-In-Time, JIT), элементов системы «канбан» (kanji + ban – визуальных карточек) по Ш. Шинго (S. Shingo) и Т. Оно (T. Ohno), оптимальной технологии производства (Optimized Production Technology, OPT) и оптимизации «узких мест» по Э. Голдратту (E. Goldratt); автоматизированная поддержка поставок (Computer-Aided Logistic Support, CALS).

История развития КИС (3/4)

- **1986-90 гг., ERP** – планирование (всех) ресурсов предприятия (Enterprise Resources Planning, ERP), в том числе человеческих (Human Resources Management, HRM) и финансовых (Financial Resources Planning, FRP).
- **1991-96 гг., SCM, CALS 2** – добавление к ERP планирования ресурсов для распределения (Distribution Resources Planning, DRP); управление цепочками поставок (Supply Chain Management, SCM), позволяющее направлять и контролировать движение материальных и информационных потоков от поставщика к потребителю, по всей цепочке; непрерывная поддержка поставок и жизненного цикла (Continuous Acquisition and Lifecycle Support , CALS 2).

История развития КИС (4/4)

- **1997-2000 гг., CSRP** – планирование ресурсов, синхронизированное с потребителями (Customer-Synchronized Resources Planning, CSRP):
интегрирование потребителей и связанных с ними подразделений с основными плановыми и производственными подразделениями, интеграция собственных информационных систем с приложениями потребителей, планирование заказов потребителей.

Структура базовых КИС (1/2)

- 1 информационная модель - представляющая собой отражение реальной информационной базы банка и описывающая все существующие информационные потоки, совокупность правил и алгоритмов функционирования информационной системы;
- 2 техническое обеспечение (суперкомпьютеры, имеющие перспективные архитектуры и технологии организации вычислительного процесса);
- 3 средства коммуникации (сетевые компьютерные технологии, технологии Internet/Intranet, технологии клиент – сервер);

Структура базовых КИС (2/2)

- 4 системное и сетевое программное обеспечение, обеспечивающее работу коммуникационных средств;
- 5 прикладное программное обеспечение, необходимое для выполнения прикладных задач в каждом подразделении банка;
- 6 средства обеспечения безопасности (разграничение доступа к ресурсам, обеспечение надежности функционирования корпоративной системы в целом)

Современное состояние базовых КИС. Тенденции (1/2)

- Основные игроки на рынке КИС - Oracle и IBM.
- ERP-системы обрастают огромным количеством модулей.
- Используются современные алгоритмы машинного обучения для построения BI-систем.

Современное состояние базовых КИС. Тенденции (2/2)

- Создание продвинутой аналитики действий пользователей.
- Построение рекомендательных моделей.
- Усиление рыночных позиций разработчиков КИС через приобретение сторонних компаний.
- Переход от клиент-серверных моделей КИС к модели с «тонким клиентом».
- Web-центризм, то есть интернет-ориентированность всех модулей КИС.

Выборочные функции распределения и моменты

Немного о временных рядах (1/2)

- Временной ряд – результат наблюдения за некоторой величиной $x(t)$, где t_1, \dots, t_n – моменты наблюдения, x_1, \dots, x_n – наблюдаемые значения величины.
- Случайный процесс – семейство $\{x(t), t \in T\}$. Если t принимает дискретные значения, то это временной ряд.

Немного о временных рядах (2/2)

- Непрерывная случайная величина ξ принимает дискретное множество значений, если наблюдается в дискретные моменты времени.
- Распределением дискретной случайной величины ξ называется множество $\Pr(\xi = x_i) = p_i$, где $\sum_{i=1}^n p_i = 1$.
- Функцией распределения называется $F(x) = \Pr(\xi \leq x)$.
- Квантилем порядка q для $F(x)$ называется такое число x_q , что $F(x_q) = q$.
- Моментом порядка r называется $m_r = \mathbf{E}\xi^r$.

Выборочная функция распределения

- Множество принимаемых значений $x(t)$ ограничено на ограниченном отрезке времени $[t_1; t_2]$, можно считать, что $x(t) \in [0; 1]$.
- Если в выборке объёма T значение x_i встретилось $n_i(T)$ раз, то относительной частотой этого события называется $\nu_i(T) = \frac{n_i(T)}{T}$.
- Соответственно выборочной функцией распределения (ВФР) называется ступенчатая неубывающая функция $F_T(x)$, определяемая по набору значений $\nu_i(T)$.
- **Теорема:** $\Pr\{\lim_{T \rightarrow \infty} \sup_x |F_T(x) - F(x)| = 0\} = 1$.

Теория тезауруса. Смысловые связи между элементами тезауруса. Роль тезауруса в поисковых системах

Теория тезауруса (1/3)

- Язык – самая развитая и совершенная знаковая система.
- Две основные функции языка – коммуникативная (реализующая общение между людьми) и кумулятивная (накопление в текстах различного рода знаний).
- В современные информационные системы включено большое количество различных словарей: толковые, словари синонимов, антонимов, омонимов и др.

Теория тезауруса (2/3)

- Слова-синонимы: «туча» и «облако», «собака» и «пёс».
- Слова-омонимы: «ключ» (от двери) и «ключ» (родник), «наряд» (одежда) и «наряд» (распоряжение).

Теория тезауруса (3/3)

- Тезаурус-1 – все слова языка, включая самые редкие, с перечнем примеров их использования.
- Тезаурус-2 – классификация знаков и денотатов в иерархические структуры.
- Важная особенность тезауруса – возможность искать незнакомые слова по их значению (денотату).

Роль тезауруса в поисковых системах

- Информационный поиск не только по заданным словам, но и по их синонимам, а так же по подмножествам значений, например, по запросу «дерево» в поисковый запрос будут также включены названия видов деревьев, такие как «дуб», «сосна», «липа».

Исследование статистического предиктора приступа эпилепсии по данным электроэнцефалограмм

Научный руководитель: д. ф.-м. н., доцент Орлов Юрий Николаевич

Введение

- **Эпилепсия** известна людям с давнейших времён. Многие поколения врачей и учёных сталкиваются с проблемой предсказания приступов эпилепсии по состоянию пациента, с проблемой дифференциации обычного состояния больного от предприпадочного состояния.
- Основным инструментом диагностики эпилепсии является ЭЭГ (электроэнцефалограмма) и «электроэнцефалография» – чтение и расшифровка результатов ЭЭГ.

Актуальность работы

- Вопросы предсказания приступа эпилепсии остро стоят перед врачами многих стран.
- Зачастую медикаментозное и хирургическое лечение противопоказаны.
- Отсутствуют стабильно действующие методы предсказания приступов.
- Течение болезни отличается от пациента к пациенту.

Цели работы

- **Цель:** построение предиктора приступа эпилепсии, который позволял бы предсказывать наступающие приступы эпилепсии за некоторое определённое время до начала приступа вне зависимости от неконсистентности поступающих данных (например, если один из электродов «отошёл» и перестал передавать показания).

Задачи работы

- 1 Исследовать поступающие на вход данные, представить их в виде независимых друг от друга временных рядов.
- 2 Попробовать найти закономерности в данных временных рядах, исследовать стационарность ряда для различных n , где n – количество интервалов, на которые разбивается ряд.
- 3 Найти значения функций $F_n(x, t_k) = \Pr(\xi < x)$ и $\rho_k = \max_x |F_n(x, t_k) - F_n(x, t_{k+1})|$, исследовать, является ли $\{\rho_k\}_1^K$ динамической системой.
- 4 На основе полученных данных попытаться построить предсказание эпилептического приступа.

Ход работы (1/5)

Входные данные представляют собой набор независимых друг от друга временных рядов данных, поступающих с различных датчиков ЭЭГ.

```
In [1]: import csv
import pandas as pd
from IPython.display import display

with open('AC pld0,8ed.csv', newline='') as csvfile:
    csvreader = csv.reader(csvfile, delimiter=',')
    raw_data = []

    for row in csvreader:
        raw_data.append(row)

    headers = raw_data.pop(0)
    df = pd.DataFrame(data=raw_data, columns=headers)
    display(df.head(10))
```

	E1	E2	FP1-A1	F7-A1	F3-A1	T3-A1	C3-A1	P3-A1	T5-A1	O1-A1	...	C4-A2	P4-A2	T6-A2	O2-A2	FPZ-A1	FZ-A2	CZ-A1	PZ-A2	E3	E4
0	0	0	-0.885	-1.149	-0.76	0.416	1.522	0.683	3.434	-9.354	...	-0.184	-2.889	-0.558	0.485	-1.25	0.676	-3.416	-1.082	-6.291	0
1	0	0	-0.633	-0.822	-0.544	0.298	1.089	0.488	2.457	-6.692	...	-0.132	-2.067	-0.399	0.347	-0.894	0.484	-2.444	-0.774	-4.517	0
2	0	0	-0.38	-0.494	-0.327	0.179	0.654	0.293	1.476	-4.021	...	-0.079	-1.242	-0.24	0.209	-0.537	0.291	-1.468	-0.465	-3.042	0
3	0	0	-0.141	-0.184	-0.121	0.067	0.243	0.109	0.549	-1.495	...	-0.029	-0.462	-0.089	0.078	-0.2	0.108	-0.546	-0.173	-1.709	0
4	0	0	0.07	0.091	0.06	-0.033	-0.121	-0.054	-0.272	0.742	...	0.015	0.229	0.044	-0.038	0.099	-0.054	0.271	0.086	-0.333	0
5	0	0	0.244	0.317	0.209	-0.115	-0.419	-0.188	-0.946	2.578	...	0.051	0.796	0.154	-0.134	0.345	-0.186	0.941	0.298	1.079	0
6	0	0	0.373	0.484	0.32	-0.175	-0.641	-0.288	-1.447	3.942	...	0.077	1.217	0.235	-0.204	0.527	-0.285	1.44	0.456	2.294	0
7	0	0	0.456	0.591	0.391	-0.214	-0.784	-0.351	-1.768	4.815	...	0.095	1.487	0.287	-0.25	0.644	-0.348	1.759	0.557	3.004	0
8	0	0	0.495	0.642	0.425	-0.233	-0.851	-0.382	-1.919	5.228	...	0.103	1.615	0.312	-0.271	0.699	-0.378	1.909	0.605	3.057	0
9	0	0	0.497	0.645	0.427	-0.234	-0.855	-0.383	-1.928	5.253	...	0.103	1.622	0.313	-0.272	0.702	-0.38	1.918	0.607	2.576	0

10 rows x 24 columns



Ход работы (2/5)

- Алгоритм написан на языке python3.6 с использованием Jupyter Notebook.
- При написании алгоритма использованы модули pandas для работы с таблицами, numpy для работы с массивами данных, matplotlib для визуализации графиков.

Ход работы (3/5)

Преобразование входных данных и представление их в качестве независимых временных рядов.

```
In [2]: data = {}

for i in range(len(headers)):
    data[headers[i]] = [float(x[i]) for x in raw_data]

for k, v in data.items():
    print('{0}: [{1}, ..., {2}]'.format(k, ', '.join([str(x) for x in v[:5]]), ', '.join([str(x) for x in v[-5:]])))

E1: [0.0, 0.0, 0.0, 0.0, 0.0, ..., 17.531, 15.763, 16.174, 15.595, 15.367]
E2: [0.0, 0.0, 0.0, 0.0, 0.0, ..., 11.688, 10.096, 10.356, 9.076, 8.827]
FP1-A1: [-0.885, -0.633, -0.38, -0.141, 0.07, ..., 14.392, 14.764, 15.572, 15.279, 12.325]
F7-A1: [-1.149, -0.822, -0.494, -0.184, 0.091, ..., -1.759, -6.438, -9.229, -9.51, -7.778]
F3-A1: [-0.76, -0.544, -0.327, -0.121, 0.06, ..., -15.131, -17.754, -17.639, -14.155, -8.111]
T3-A1: [0.416, 0.298, 0.179, 0.067, -0.033, ..., -5.981, -10.59, -12.733, -11.928, -8.869]
C3-A1: [1.522, 1.089, 0.654, 0.243, -0.121, ..., -16.84, -19.095, -18.488, -14.263, -7.215]
P3-A1: [0.683, 0.488, 0.293, 0.109, -0.054, ..., -13.175, -15.67, -15.7, -12.36, -6.187]
T5-A1: [3.434, 2.457, 1.476, 0.549, -0.272, ..., -8.069, -13.214, -16.32, -15.852, -11.797]
O1-A1: [-9.354, -6.692, -4.021, -1.495, 0.742, ..., 0.883, -2.915, -4.926, -4.431, -1.888]
FP2-A2: [2.606, 1.864, 1.12, 0.416, -0.207, ..., 12.333, 11.815, 11.868, 11.432, 9.224]
F8-A2: [0.115, 0.082, 0.05, 0.018, -0.009, ..., 2.625, -0.207, -1.504, -1.298, -0.483]
F4-A2: [4.661, 3.335, 2.004, 0.745, -0.37, ..., -1.373, -1.879, -0.761, 1.467, 3.468]
T4-A2: [1.478, 1.058, 0.635, 0.236, -0.117, ..., 4.766, -0.384, -3.499, -3.971, -2.557]
C4-A2: [-0.184, -0.132, -0.079, -0.029, 0.015, ..., -7.513, -9.593, -9.453, -6.758, -2.431]
P4-A2: [-2.889, -2.067, -1.242, -0.462, 0.229, ..., -5.111, -7.654, -8.14, -6.099, -2.299]
T6-A2: [-0.558, -0.399, -0.24, -0.089, 0.044, ..., 3.68, 0.843, -0.289, 0.336, 1.84]
O2-A2: [0.485, 0.347, 0.209, 0.078, -0.038, ..., 3.687, 0.369, -1.38, -1.007, 0.977]
FP8-A1: [-1.25, -0.894, -0.537, -0.2, 0.099, ..., 13.592, 13.807, 14.604, 14.515, 11.982]
F2-A2: [0.676, 0.484, 0.291, 0.108, -0.054, ..., -15.806, -14.745, -10.948, -4.95, 1.623]
C2-A1: [-3.416, -2.444, -1.468, -0.546, 0.271, ..., -16.892, -18.125, -16.329, -11.03, -3.34]
P2-A2: [-1.082, -0.774, -0.465, -0.173, 0.086, ..., -4.352, -6.783, -7.2, -4.956, -0.649]
E3: [-6.291, -4.517, -3.042, -1.709, -0.333, ..., 153.287, 119.408, -58.128, 199.121, 770.502]
E4: [0.0, 0.0, 0.0, 0.0, 0.0, ..., 1.371, 0.495, -0.358, -1.077, -1.56]
```

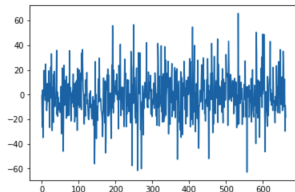
Ход работы (4/5)

Представление одного из временных рядов (на графике каждое 1000-е показание).

```
In [3]: import matplotlib.pyplot as plt
import numpy as np
%matplotlib inline

plt_data = data['FP1-A1'][::1000]
print('Max: {}'.format(max(plt_data)))
print('Min: {}'.format(min(plt_data)))
plt.plot(plt_data)
plt.show()
```

Max: 65.781.
Min: -62.993.



Ход работы (5/5)

- Разбиение временных рядов на n равных отрезков (для различных $n \in [10^3; 10^4]$, построение выборочных функций распределения (ВФР) $F_n(x, t_k)$).
- Нахождение значения функции $\rho_k = \max_x |F_n(x, t_k) - F_n(x, t_{k+1})|$ для того, чтобы понять, стационарен ряд в некотором приближении или нет.

Дальнейшие планы:

Опираясь на полученные результаты, попытаться построить предиктор эпилептического приступа.

Спасибо за внимание!