

Машинное обучение

Кластеризация и визуализация

Калтович Артём
KaltovichArtyom@tut.by

3 марта 2018 г.

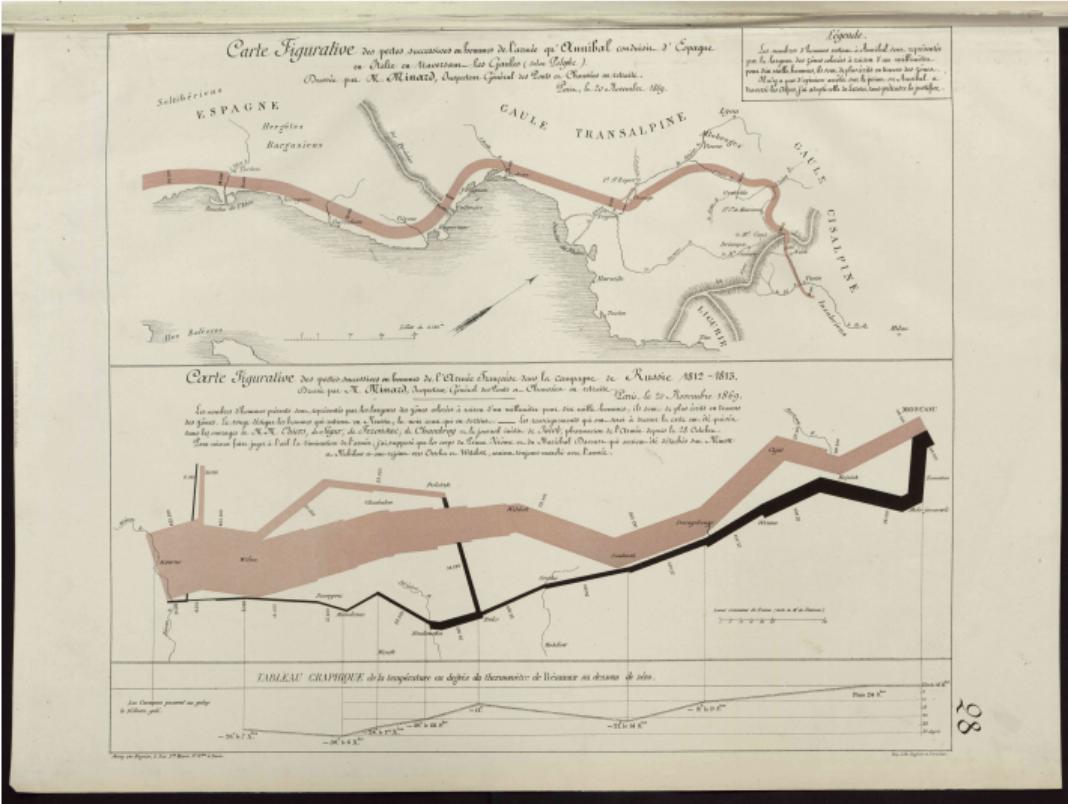
Визуализация

Определение

Визуализация (от лат. *visualis*, «зрительный», англ. *Visualization*) — общее название приёмов представления числовой информации или физического явления в виде, удобном для зрительного наблюдения и анализа. [1].

Визуализация

Пример



Визуализация

Пример

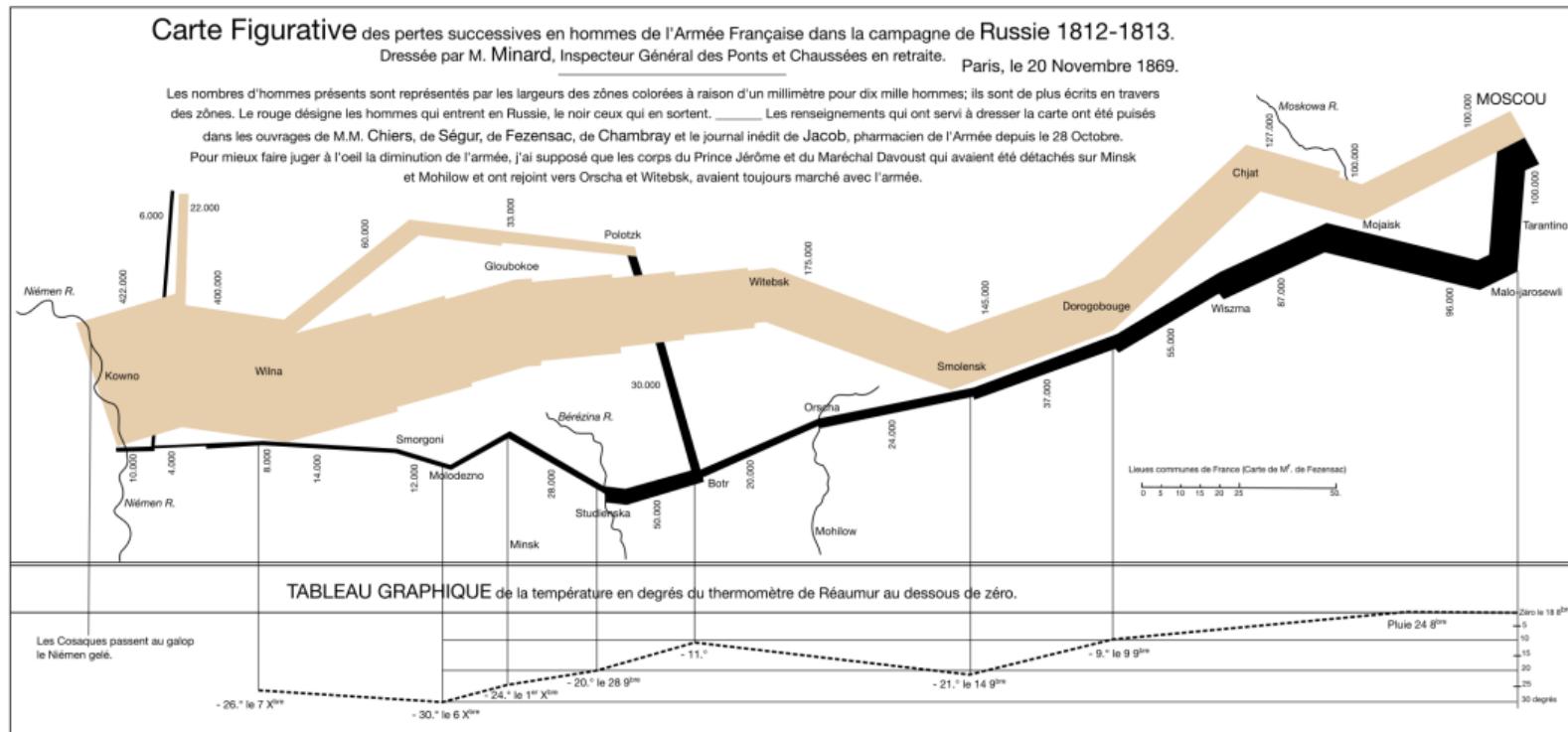
Шарль Жозеф Минар (27 марта 1781, Дижон — 24 октября 1870, Бордо).

29 ноября 1869 года — графическая визуализация вторжения Наполеона Бонапарта в Россию в 1812 году.

20 ноября 1869 года — карта, показывающая перемещение войск Ганнибала из Иберии (Испании) в Италию во время Второй Пунической войны.

Визуализация

Пример



Визуализация

Пример

Почему мы визуализируем?

Визуализация

Почему мы визуализируем?

Визуализация

Почему мы визуализируем?

Balance table: unw									16:28 Monday, January 6, 2014	2	
Obs	row_name	tx_mn	tx_sd	ct_mn	ct_sd	std_eff_sz	stat	p	ks	ks_pval	table_name
1	unw.age	25.816	7.155	28.03	10.787	-0.309	-2.994	0.003	0.158	0.003	unw
2	unw.educ	10.346	2.011	10.235	2.855	0.055	0.547	0.584	0.111	0.074	unw
3	unw.black	0.843	0.365	0.203	0.403	1.757	19.371	0	0.64	0	unw
4	unw.hispan	0.059	0.237	0.142	0.35	-0.349	-3.413	0.001	0.083	0.317	unw
5	unw.nodegree	0.708	0.456	0.597	0.491	0.244	2.716	0.007	0.111	0.074	unw
6	unw.married	0.189	0.393	0.513	0.5	-0.824	-8.607	0	0.324	0	unw
7	unw.re74	2095.574	4886.62	5619.237	6788.751	-0.721	-7.254	0	0.447	0	unw
8	unw.re75	1532.055	3219.251	2466.484	3291.996	-0.29	-3.282	0.001	0.288	0	unw

Balance table: ks.max.ATT									16:44 Monday, January 6, 2014	3	
Obs	row_name	tx_mn	tx_sd	ct_mn	ct_sd	std_eff_sz	stat	p	ks	ks_pval	table_name
17	ks.max.ATT.age	25.816	7.155	25.764	7.408	0.007	0.055	0.956	0.107	0.919	ks.max.ATT
18	ks.max.ATT.educ	10.346	2.011	10.572	2.14	-0.113	-0.712	0.477	0.107	0.919	ks.max.ATT
19	ks.max.ATT.black	0.843	0.365	0.835	0.371	0.022	0.187	0.852	0.008	1	ks.max.ATT
20	ks.max.ATT.hispan	0.059	0.237	0.043	0.203	0.069	0.779	0.436	0.016	1	ks.max.ATT
21	ks.max.ATT.nodegree	0.708	0.456	0.601	0.49	0.235	1.1	0.272	0.107	0.919	ks.max.ATT
22	ks.max.ATT.married	0.189	0.393	0.199	0.4	-0.024	-0.169	0.866	0.01	1	ks.max.ATT
23	ks.max.ATT.re74	2095.574	4886.62	1673.666	3944.6	0.086	0.8	0.424	0.054	1	ks.max.ATT
24	ks.max.ATT.re75	1532.055	3219.251	1257.242	2674.922	0.085	0.722	0.471	0.094	0.971	ks.max.ATT

Balance table: es.mean.ATT									16:44 Monday, January 6, 2014	4	
Obs	row_name	tx_mn	tx_sd	ct_mn	ct_sd	std_eff_sz	stat	p	ks	ks_pval	table_name
9	es.mean.ATT.age	25.816	7.155	25.802	7.279	0.002	0.015	0.988	0.122	0.892	es.mean.ATT
10	es.mean.ATT.educ	10.346	2.011	10.573	2.089	-0.113	-0.706	0.48	0.099	0.977	es.mean.ATT
11	es.mean.ATT.black	0.843	0.365	0.842	0.365	0.003	0.027	0.978	0.001	1	es.mean.ATT
12	es.mean.ATT.hispan	0.059	0.237	0.042	0.202	0.072	0.804	0.421	0.017	1	es.mean.ATT
13	es.mean.ATT.nodegree	0.708	0.456	0.609	0.489	0.218	0.967	0.334	0.099	0.977	es.mean.ATT
14	es.mean.ATT.married	0.189	0.393	0.189	0.392	0.002	0.012	0.99	0.001	1	es.mean.ATT
15	es.mean.ATT.re74	2095.574	4886.62	1556.93	3801.566	0.11	1.027	0.305	0.066	1	es.mean.ATT
16	es.mean.ATT.re75	1532.055	3219.251	1211.575	2647.615	0.1	0.833	0.405	0.103	0.969	es.mean.ATT

Визуализация

Попробуем сами



Рональд Фишер в 1936 году продемонстрировал работу разработанного им метода анализа.

Данные были собраны американским ботаником Эдгаром Андерсоном.

Визуализация

Попробуем сами

Признаки:

- ▶ Длина наружной доли околоцветника (англ. sepal length);
- ▶ Ширина наружной доли околоцветника (англ. sepal width);
- ▶ Длина внутренней доли околоцветника (англ. petal length);
- ▶ Ширина внутренней доли околоцветника (англ. petal width).

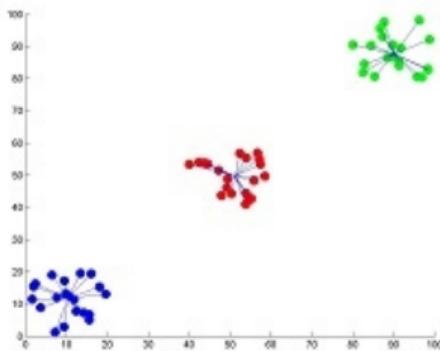
Визуализация

Попробуем сами

Примеры кода

Кластеризация

Определение



Кластерный анализ (англ. *cluster analysis*) — многомерная статистическая процедура, выполняющая сбор данных, содержащих информацию о выборке объектов, и затем упорядочивающая объекты в сравнительно однородные группы.[2]

Визуализация

Набор ирисов Фишера



► Ирис щетинистый
(лат. *Iris setosa*)



► *Iris versicolor*



► *Iris virginica*

Визуализация

Набор ирисов Фишера

Примеры кода

Визуализация

Набор ирисов Фишера

Задание:

- ▶ Выбрать любые две пары признаков
- ▶ Отобразить на одном рисунке две зависимости
- ▶ Подписать график и оси
- ▶ Добавить легенду.

Кластеризация

Методы

Методы:

- ▶ К-средних (k-means)

Машинное обучение

Определение

Машинное обучение (англ. machine learning, ML) — класс методов искусственного интеллекта, характерной чертой которых является не прямое решение задачи, а обучение в процессе применения решений множества сходных задач. Для построения таких методов используются средства математической статистики, численных методов, методов оптимизации, теории вероятностей, теории графов, различные техники работы с данными в цифровой форме[3].

Алгоритм A обучается с эффективностью E над данными D , если при росте мощности $|D|$, E проявляет тенденцию к увеличению.

Список литературы