

Социальные медиа

Дескриптивный и структурный анализ. Подходы и Методы.

Макаренко А.В.

avm.science@mail.ru

**Научно-исследовательская группа
«Конструктивная Кибернетика»
www.rdcn.ru**

**Институт проблем Управления
Российская академия наук
Москва, Россия**

02 декабря 2015 г.

□ План мини-курса «Социальные медиа»

Мини-курс «Социальные медиа»:

Лекция х.1. Большие Данные. Цели, задачи и парадигма анализа.

Лекция х.2. Дескриптивный и структурный анализ. Подходы и Методы.

Лекция х.3. Прикладные задачи анализа. Проблематика.

□ План II лекции

1. Анализ социальных медиа в контексте Data Science.
2. Основы экосистемы R.
3. Дескриптивная статистика. Заполнение пропусков.
4. Регрессионные модели. Понижение размерности данных.
5. Графовые модели. Топология данных.
6. Вторичная формализация данных. Поиск информативных признаков.
7. Машинное обучение. Структурные модели. Вскрытие смысла.
8. Когнитивная визуализация. Бизнес-инфографика.
9. Задания для самостоятельной проработки.

□ Анализ социальных медиа в контексте Data Science.

Что такое Data Science?

- **Data Science** – раздел информатики, изучающий проблемы анализа, обработки и представления данных в цифровой форме. Объединяет методы по обработке данных в условиях больших объёмов и высокого уровня параллелизма, статистические методы, методы интеллектуального анализа данных и приложения искусственного интеллекта для работы с данными, а также методы проектирования и разработки баз данных. [[wikipedia](#)]
- **Data Mining** – собирательное название, используемое для обозначения совокупности методов обнаружения в данных ранее неизвестных, нетривиальных, практически полезных и доступных интерпретации знаний, необходимых для принятия решений в различных сферах человеческой деятельности. [[wikipedia](#)]
- **Машинное обучение** – обширный подраздел искусственного интеллекта, математическая дисциплина, использующая разделы математической статистики, численных методов оптимизации, теории вероятностей, дискретного анализа выделяющая знания из данных. В настоящий момент упор делается на обучении по прецедентам, или индуктивном обучении, основаном на выявлении закономерностей в эмпирических данных. [[wikipedia](#)]

□ Анализ социальных медиа в контексте Data Science.

Базовая Триада задач Data Science



□ Анализ социальных медиа в контексте Data Science.

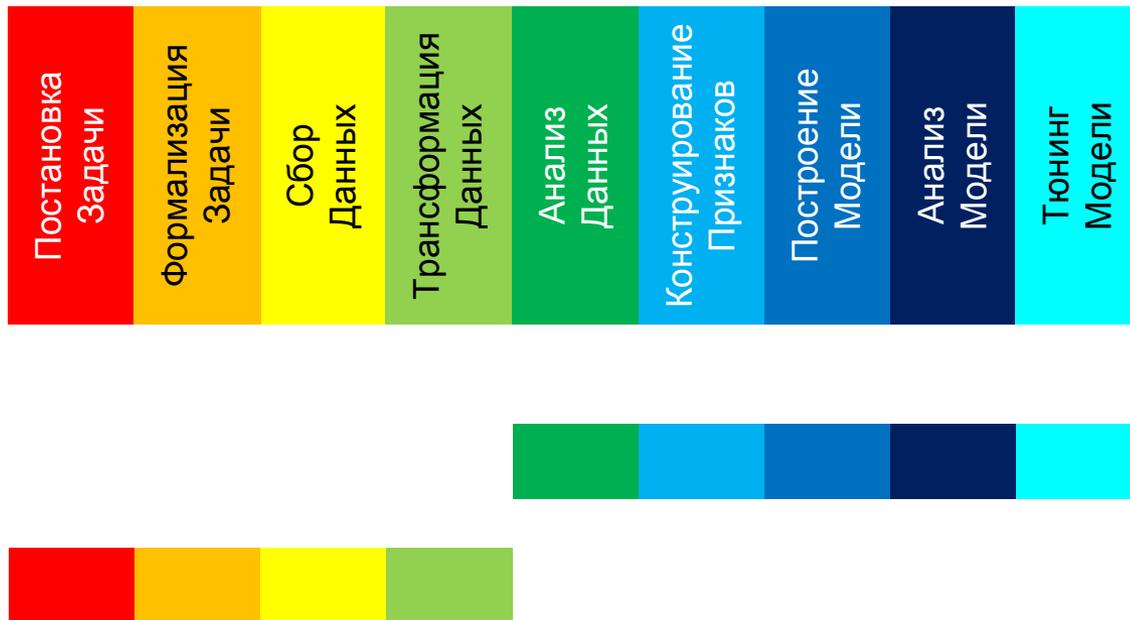
Конвейер Data Science



- Источник представления:
 - Курсы (Coursera, ШАД, ...).
 - Соревнования (Kaggle, ...) и т.п.
- Особенности:
 - Работа с "рафинированными" данными (как правило).
 - Работа вне контекста бизнес-процессов и промышленных инфосистем.

□ Анализ социальных медиа в контексте Data Science.

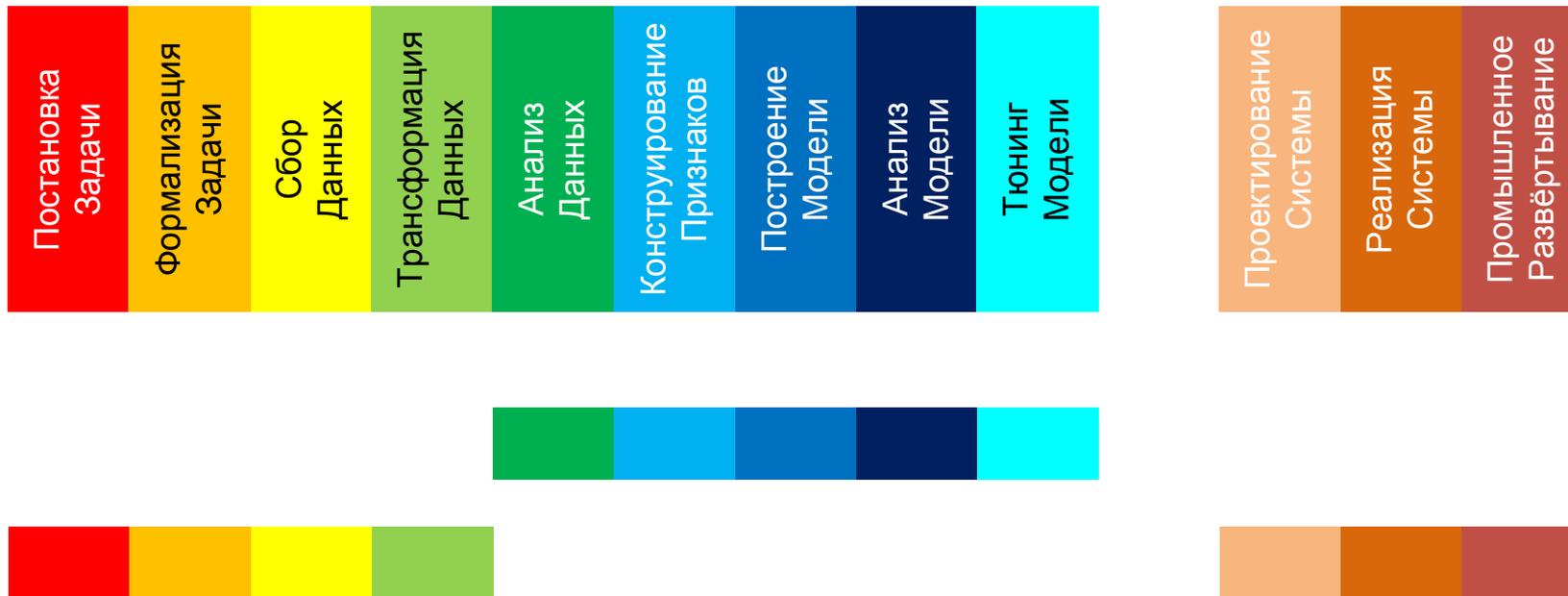
Конвейер Data Science



- Источник представления:
 - Поставщики данных (Bloomberg, ...) и т.п.
- Особенности:
 - Узость предметной области; Дорогие данные.
 - Нестыковки в постановках задач, и т.п.

□ Анализ социальных медиа в контексте Data Science.

Конвейер Data Science



- Источник представления:
 - Разработчики программного обеспечения ("[тыжпрограммист](#)", ...) и т.п.
- Особенности:
 - Владение стеком технологий в отрыве от алгоритмов и бизнес-процессов.

□ Анализ социальных медиа в контексте Data Science.

Конвейер Data Science



- Источник представления:
 - Разработчики полного цикла.
- Особенности:
 - Их мало (толковых), Они загружены, Они дорогие.

□ Анализ социальных медиа в контексте Data Science.

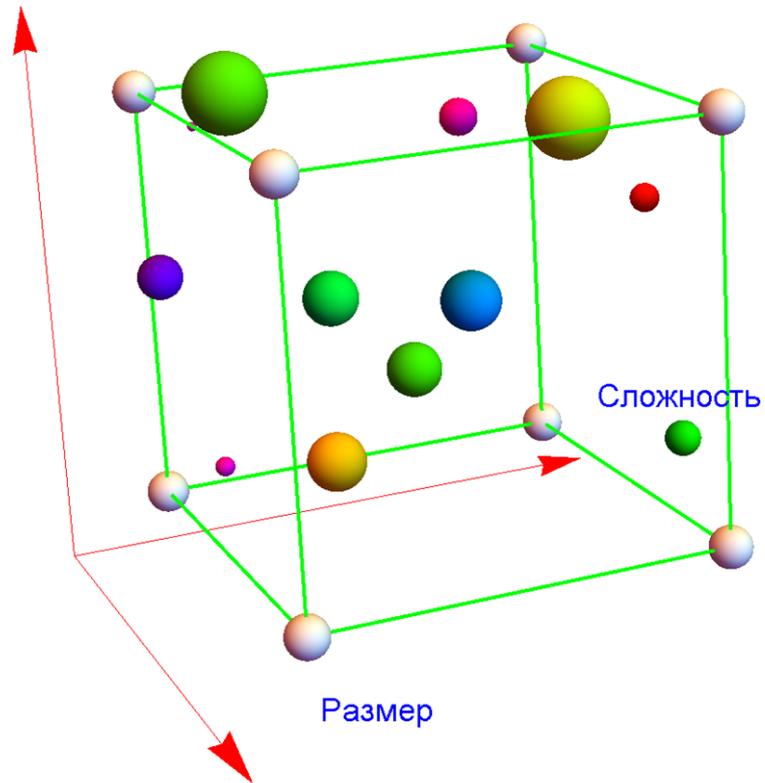
Конвейер Data Science



□ Анализ социальных медиа в контексте Data Science.

Классификация задач Data Science

Гетерогенность



□ Анализ социальных медиа в контексте Data Science.

Классификация задач Data Science

- **Размер задачи** – определяется количеством операций и/или объёмом данных (на уровне размерности вектора и количества объектов), требуемых для решения задачи.
- **Сложность задачи** – зависит от существования «вычислимого» алгоритма решения задачи и его переносимости на компьютерные платформы.
- **Гетерогенность задачи** – определяется количеством разделов математики требующихся для решения задачи и разнородностью входных данных.

Важный контекст задач:

- *Медианные* – типовые потребности большинства Аналитиков.
- *Экстремальные* – ~~за гранью добра добра и зла~~ редкие и специфичные задачи.

□ Основы экосистемы R.

Основные операторы языка R

- Оператор присваивания `<-`

```
> n <- 16
```

```
> n
```

```
[1] 16
```

```
>
```

```
> m <- 2*(3-4)
```

```
> m
```

```
[1] -2
```

```
>
```

- Вычисление без присваивания

```
> 2*(3-4)
```

```
[1] -2
```

```
>
```

□ Основы экосистемы R.

Основные операторы языка R

- Отображение имён объектов `ls()`

```
> ls()  
[1] "m" "n"  
>
```

- Отображение характеристик объектов `str()`

```
> str(m)  
  num -2  
>
```

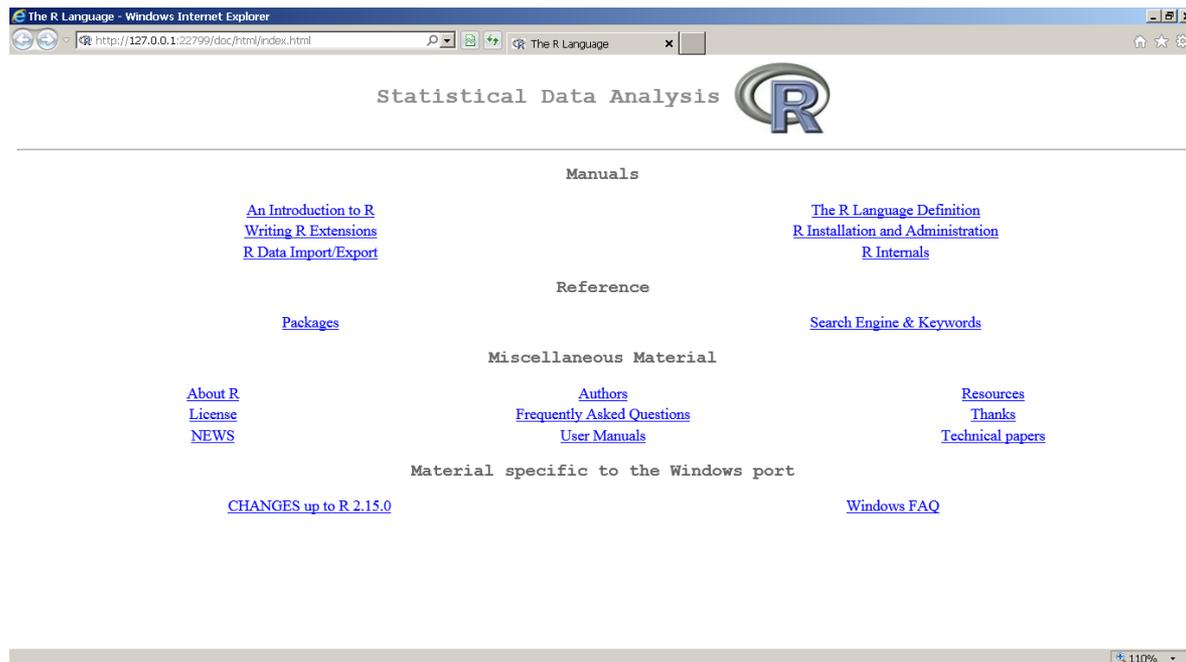
- Удаление объекта из памяти `rm()`

```
> rm(n)  
> ls()  
[1] "m"  
>
```

❑ Основы экосистемы R.

Основные операторы языка R

- Вызов справки в формате HTML `help.start()`



- Контекстная справка `help()` или ?

```
> ?ls
ls {base}
List Objects
```

...

□ Основы экосистемы R.

Основные типы и структуры данных в языке R

- Типы данных

numeric (числовой) [integer, double] : 5 , -12.6 , 5.1E-10

complex (комплексный) : 7.0+1i*5.5

character (символьный) : 'a' , "data"

logical (логический) : FALSE , TRUE

factor (категориальный) : Factor w/ 3 levels "b","g","r" : 3 2 1

- Базовые характеристики объектов

- Тип данных `typeof()`
- Размер `length()`
- Размерность `dim()`
- Имя `names()`

□ Основы экосистемы R.

Основные типы и структуры данных в языке R

- Структуры данных

Объект	Размерность объекта	Монотип данных	Регулярность структуры
vector	1 [length]	Да	Да
matrix	2 [rows x cols]	Да	Да
array	N [axis1 x axis2 x ...]	Да	Да
list	1 [length]	Нет	Нет
ts	N [length x axis1 x axis2 x ...]	Да	Да
data.frame	2 [rows x cols]	Нет	Да

- Проверка типа объекта

```
> is.numeric(m)
[1] TRUE
>
```

- Проверка структуры объекта

```
> is.list(m)
[1] FALSE
>
```

□ Основы экосистемы R.

Генерация данных в системе R

- Регулярная последовательность целых чисел :

```
> x <- 1:10
> x
[1] 1 2 3 4 5 6 7 8 9 10
>
```

- Регулярная последовательность действительных чисел `seq()`

```
> y <- seq(1, 5, 0.5)
> y
[1] 1.0 1.5 2.0 2.5 3.0 3.5 4.0 4.5 5.0
>
```

- Вектор идентичных элементов `rep()`

```
> z <- rep("z", 10)
> z
[1] "z" "z" "z" "z" "z" "z" "z" "z" "z" "z"
>
```

□ Основы экосистемы R.

Генерация данных в системе R

- Последовательность заданных элементов `c()`

```
> w <- c(1, 2.3, "str")
```

```
> w
```

```
[1] "1"    "2.3"  "str"
```

```
>
```

```
> is.vector(w)
```

```
[1] TRUE
```

```
>
```

- Случайные последовательности `r...()`

- Нормальное распределение

```
> v <- rnorm(5, mean = 1, sd = 0.1)
```

```
> v
```

```
[1] 1.2424296 1.0022309 1.2370456 0.9749742 0.9555599
```

```
>
```

❑ Основы экосистемы R.

Загрузка данных в систему R

- Чтение csv файлов `read.csv()`, `read.csv2()`

```
> dt1 <- read.csv("file_name", header = TRUE, sep = ",",  
quote = "\"", dec = ".", ...)  
> dt2 <- read.csv2("file_name", header = TRUE, sep = ";",  
quote = "\"", dec = ",", ...)  
>
```

- Чтение бинарных файлов `readBin()`

```
> fs <- file("file_name", "rb")  
> dt3 <- readBin(fs, integer(), n = 4)  
> close(fs)  
>
```

- Загрузка содержимого web страниц `getURL()`

```
> require(RCurl)  
> webpage <- getURL("http://www.rdcn.ru/")  
> is.character(webpage)  
[1] TRUE  
>
```

□ Основы экосистемы R.

Доступ к определённому элементу объекта в системе R

- По индексу []

```
> p <- 1:5  
> p[2:4]  
[1] 2 3 4  
>
```

- По значению []

```
> p[p > 3]  
[1] 4 5  
>
```

- По имени []

```
> names(p) <- c("a", "b", "c", "d", "e")  
> p["b"]  
b  
2  
>
```

□ Основы экосистемы R.

Обработка данных в системе R

- Цикл `for()`

```
> j = 0;
> for(i in 1:5) {
+   j = j+ i^2
+ }
> j
[1] 55
>
```

- Условный переход `ifelse()`

```
> x <- 1:10
> ifelse(x < 5 | x > 8, x, 0)
[1] 1 2 3 4 0 0 0 0 9 10
>
```

□ Основы экосистемы R.

Обработка данных в системе R

- Векторизованные функции ... ()

```
> x <- 1:5
> y <- sin(x)
> y
[1] 0.8414710 0.9092974 0.1411200 -0.7568025 -0.9589243
>
```

- Функциональный подход `apply()`

```
> x <- matrix(1:10, 2, 5)
> y <- apply(x, 2, min)
> y
[1] 1 3 5 7 9
>
```

- Определение новых функций `function()`

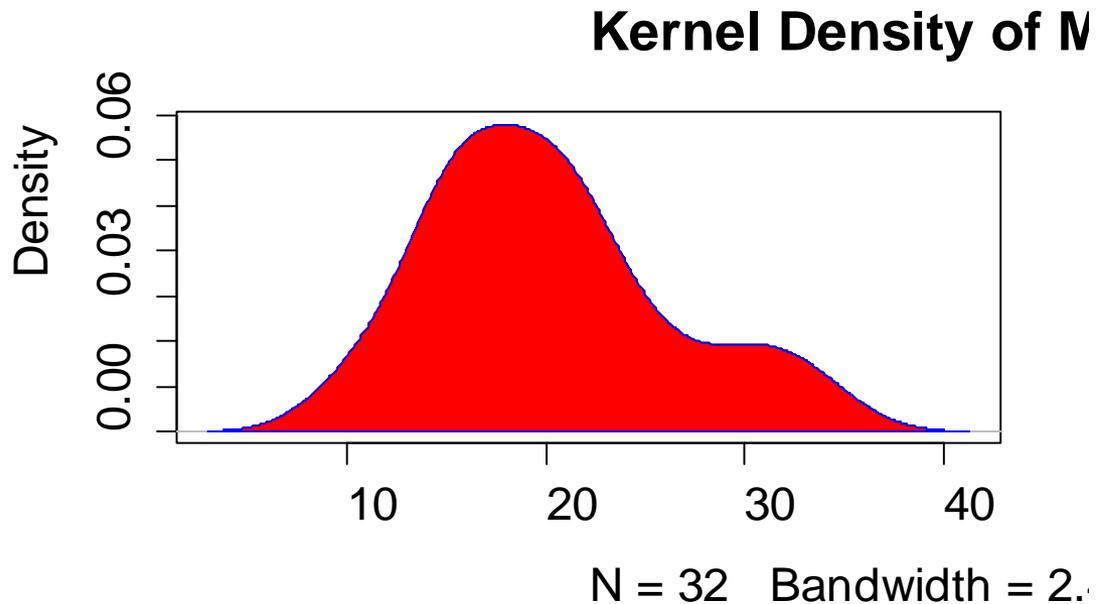
```
> scx <- function(x){sin(x)+cos(x)}
> scx(1:5)
[1] 1.3817733 0.4931506 -0.8488725 -1.4104461 -0.6752621
>
```

❑ Основы экосистемы R.

Визуализация в системе R

- Функция `plot()`

```
> d <- density(mtcars$mpg)
> plot(d, main = "Kernel Density of Miles Per Gallon")
> polygon(d, col = "red", border = "blue")
>
```



❑ Основы экосистемы R.

Расширение возможностей системы R

- Функция **install.packages()**

```
> install.packages("caret")
trying URL 'https://cran.rstudio.com/bin/windows/contrib/3.2/
caret_6.0-62.zip'
Content type 'application/zip' length 4553073 bytes (4.3 MB)
downloaded 4.3 MB
```

```
package 'caret' successfully unpacked and MD5 sums checked
```

```
The downloaded binary packages are in
  C:\TEMP\Rtmpw7U2gR\downloaded_packages
>
```

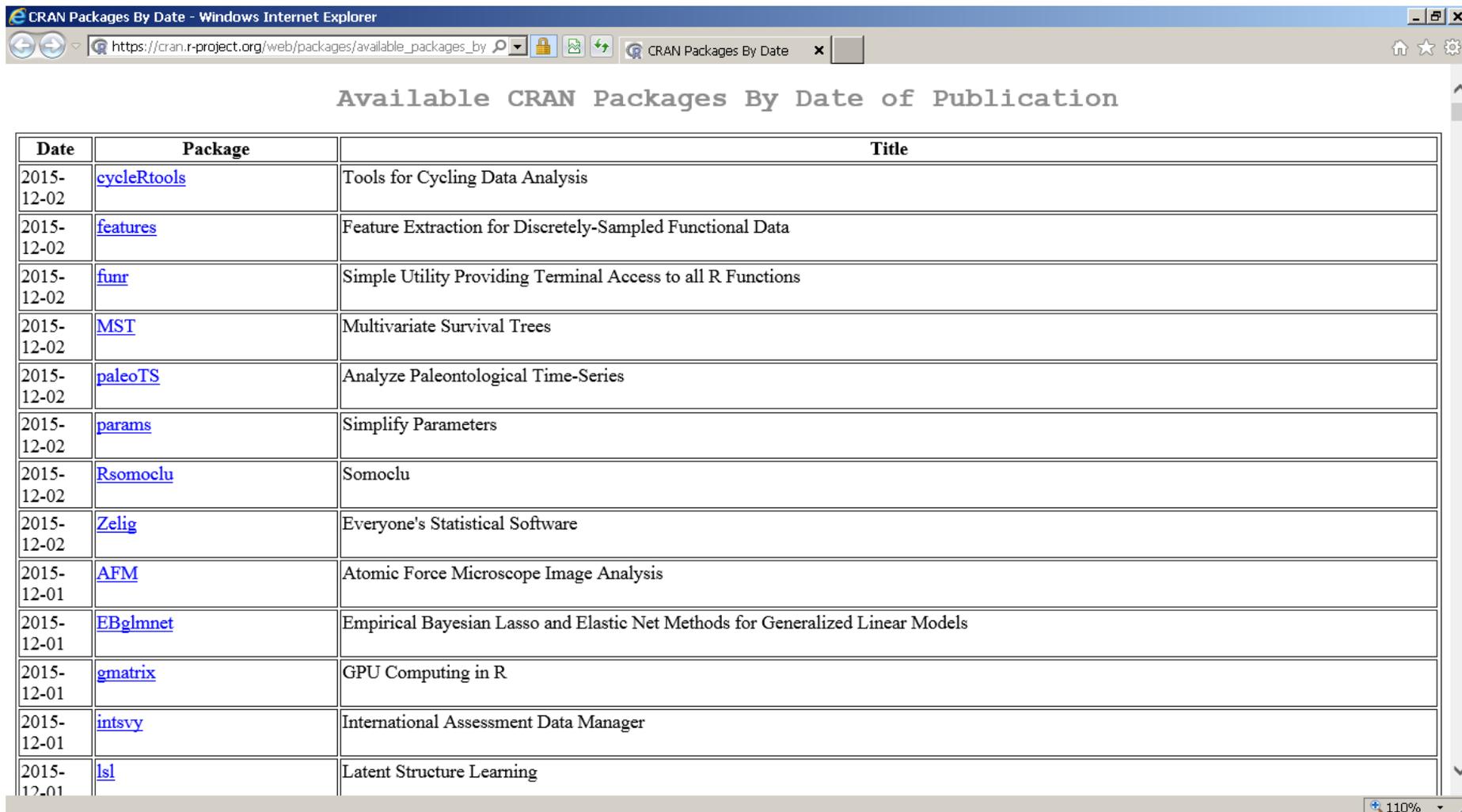
- Функция **require()**

```
> require("caret")
Загрузка требуемого пакета: caret
Загрузка требуемого пакета: lattice
Загрузка требуемого пакета: ggplot2
>
```

❑ Основы экосистемы R.

Расширение возможностей системы R

Репозиторий CRAN: 7 554 пакета.



Date	Package	Title
2015-12-02	cycleRtools	Tools for Cycling Data Analysis
2015-12-02	features	Feature Extraction for Discretely-Sampled Functional Data
2015-12-02	funr	Simple Utility Providing Terminal Access to all R Functions
2015-12-02	MST	Multivariate Survival Trees
2015-12-02	paleoTS	Analyze Paleontological Time-Series
2015-12-02	params	Simplify Parameters
2015-12-02	Rsomoclu	Somoclu
2015-12-02	Zelig	Everyone's Statistical Software
2015-12-01	AFM	Atomic Force Microscope Image Analysis
2015-12-01	EBglmnet	Empirical Bayesian Lasso and Elastic Net Methods for Generalized Linear Models
2015-12-01	gmatrix	GPU Computing in R
2015-12-01	intsvy	International Assessment Data Manager
2015-12-01	lsl	Latent Structure Learning

https://cran.r-project.org/web/packages/available_packages_by_date.html

❑ Дескриптивная статистика. Заполнение пропусков.

Пропуски и аномальные значения в данных

#	P1	P2	P3	P4	P5	P6
1		b	5	8.1	g	4
2	5	a	3	9.4	c	
3	9	c		5.7	k	1
4	1	b	4	1.3	k	
5		g	9	6.8	d	3
6	8	d	1	7.3		5
7	9	c		2.5	b	2
8	4		5	9.8	a	
9	6	a	4	6.2	0	1
10	8		3	3.4		5
11	7	d	1		f	3
12	11		9	2.6	k	6

Особенности работы с Социальными Медиа:

- Слабая структурированность данных
- Пропуски в данных
- Аномальные значения в данных

□ **Дескриптивная статистика. Заполнение пропусков.**

Пропуски и аномальные значения в данных

#	P1	P2	P3	P4	P5	P6
1		b	5	8.1	g	4
2	5	a	3	9.4	c	
3	9	c		5.7	k	1
4	1	b	4	1.3	k	
5		g	9	6.8	d	3
6	8	d	1	7.3		5
7	9	c		2.5	b	2
8	4		5	9.8	a	
9	6	a	4	6.2	0	1
10	8		3	3.4		5
11	7	d	1		f	3
12	11		9	2.6	k	6

Причины пропусков в данных:

- Отсутствие данных
- Запрет на доступ к данным
- Проблемы с ПО считывания данных

Варианты борьбы с пропусками в данных:

- Фильтрация набора данных
- Заполнение медианными значениями
- Заполнение на основе модельной ФРВ
- Заполнение на основе выборочной ФРВ

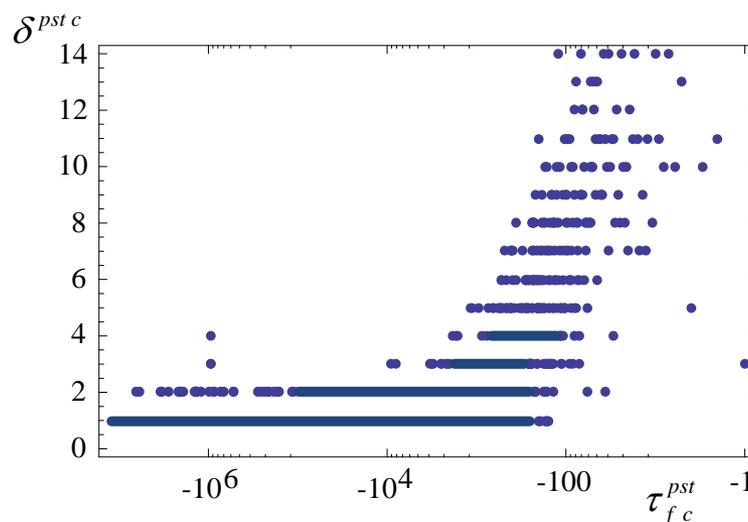
□ **Дескриптивная статистика. Заполнение пропусков.**

Пропуски и аномальные значения в данных

#	P1	P2	P3	P4	P5	P6
1		b	5	8.1	g	4
2	5	a	3	9.4	c	
3	9	c		5.7	k	1
4	1	b	4	1.3	k	
5		g	9	6.8	d	3
6	8	d	1	7.3		5
7	9	c		2.5	b	2
8	4		5	9.8	a	
9	6	a	4	6.2	0	1
10	8		3	3.4		5
11	7	d	1		f	3
12	11		9	2.6	k	6

Причины аномальных значений в данных:

- Искажение на уровне пользователя
- Искажение на уровне площадки СМ
- Проблемы с ПО считывания данных
- Неверные априорные представления



□ **Дескриптивная статистика. Заполнение пропусков.**

Пропуски и аномальные значения в данных

- Пропущенные данные **NA**

```
> h <- c(8, 10, NA, NA, 8, NA, 8)
```

```
> h
```

```
[1] 8 10 NA NA 8 NA 8
```

```
>
```

```
> mean(h)
```

```
[1] NA
```

```
>
```

```
> mean(h, na.rm = TRUE)
```

```
[1] 8.5
```

```
>
```

```
> h[is.na(h)] <- mean(h, na.rm = TRUE)
```

```
> h
```

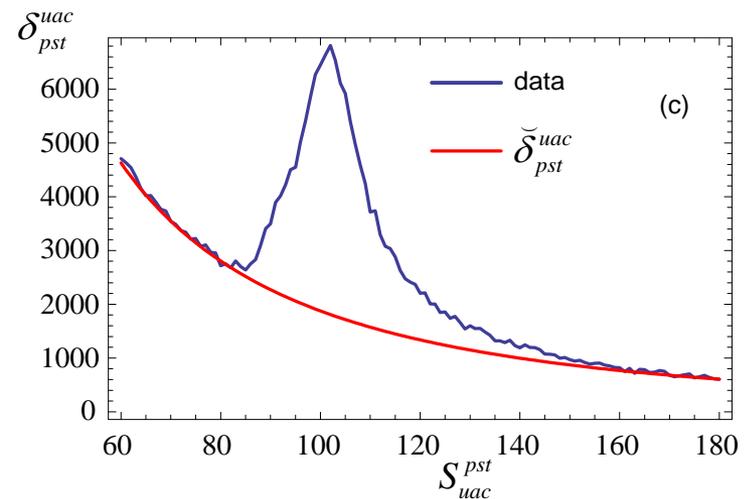
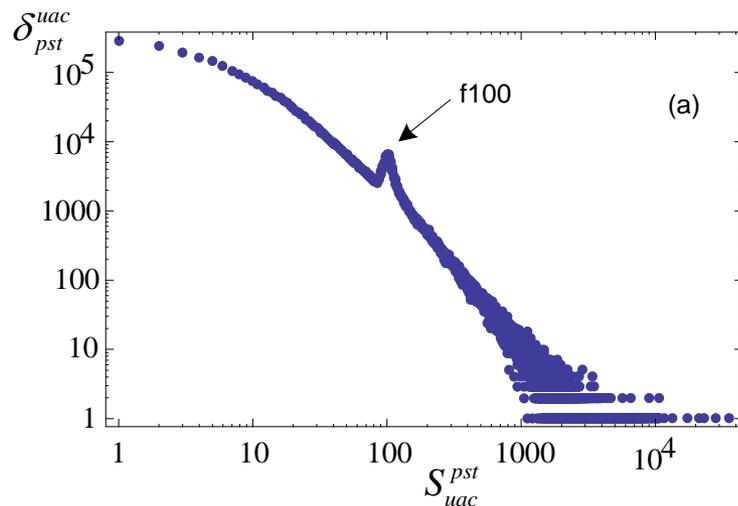
```
[1] 8.0 10.0 8.5 8.5 8.0 8.5 8.0
```

```
>
```

□ **Дескриптивная статистика. Заполнение пропусков.**

Разведочный анализ. Дескриптивная статистика.

Прежде чем начинать заполнение пропусков и исправление аномальных значений, необходимо удостовериться в том, что пропуски – это пропуски, а аномалии – это ошибки в данных. Возможно придётся скорректировать свои априорные модели и представления, а данные – оставить как есть.



Читать:

С.А. Айвазян и др. Прикладная статистика: Основы моделирования и первичная обработка данных, М.: Финансы и статистика, 1983.

□ Регрессионные модели. Понижение размерности данных.

Регрессионные модели и анализ.

$$y = g(x) + \xi$$

g – функция регрессионной модели.

ξ – случайная величина с нулевым мат. ожиданием.

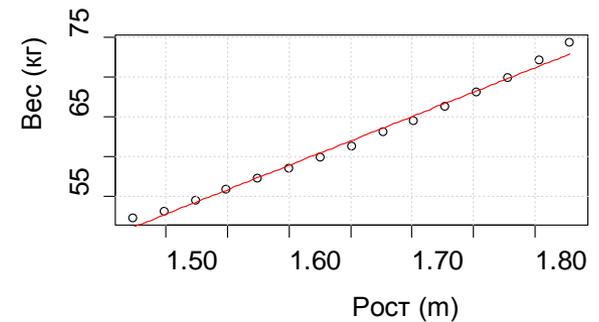
$F(\xi)$ – ФРВ есть гипотеза порождения данных.

Подробнее: <http://www.machinelearning.ru/wiki/index.php> (Регрессионный анализ)

□ Регрессионные модели. Понижение размерности данных.

Регрессионные модели и анализ.

```
> women.metr <- women
> women.metr$height <- 0.0254*women.metr$height
> women.metr$weight <- 0.45359237*women.metr$weight
>
> lm.women<-lm(formula = weight ~ height, data = women.metr)
> lm.women$coefficients
(Intercept)      height
  -39.69689      61.60999
>
> b0 <- lm.women$coefficient[1]
> b1 <- lm.women$coefficient[2]
> x1 <- min(women.metr$height)
> x2 <- max(women.metr$height)
> x <- seq(from = x1, to = x2, length.out = 100)
> y <- b0 + b1*x
>
> plot(women.metr$height, women.metr$weight, main = "",
xlab = "Рост (m)", ylab = "Вес (кг)")
> grid()
> lines(x, y, col = "red")
>
```



Пример из книги: Шипунов А.Б. и др. Наглядная статистика. Используем R!

□ Регрессионные модели. Понижение размерности данных.

Понижение размерности данных.

Типичная ситуация при анализе социальных медиа – это весьма высокая размерность входного вектора признаков (вполне возможны варианты с размерностью более 1000).

Проблемы:

- Затраты памяти и времени при обработке данных
- Сложность визуализации данных
- Сложность построения априорных моделей систем и процессов

□ Регрессионные модели. Понижение размерности данных.

Понижение размерности данных.

Решение: понижение размерности входного вектора признаков с минимальной потерей информации в аспекте решаемой задачи.

Основные подходы:

- Линейный и нелинейных методы главных компонент (PCA)
- Нелинейные методы главных многообразий

Читать:

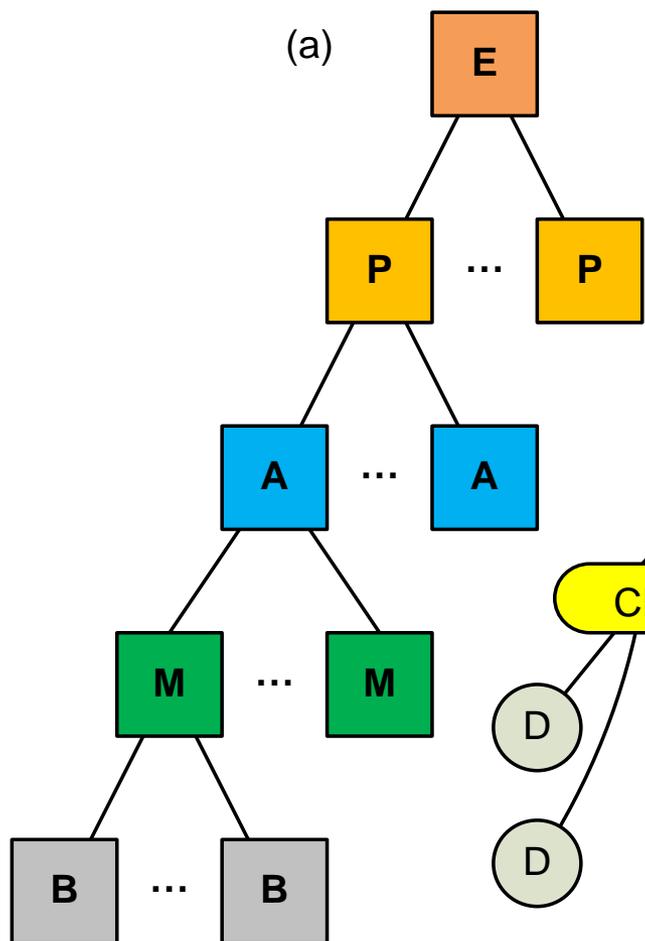
A.N. Gorban, B. Kegl, D. Wunsch, A. Zinovyev (Eds.), *Principal Manifolds for Data Visualisation and Dimension Reduction*, LNCSE 58, Springer: Berlin – Heidelberg – New York, 2007.

С.А. Айвазян и др. *Прикладная статистика: Классификации и снижение размерности*, М.: Финансы и статистика, 1989.

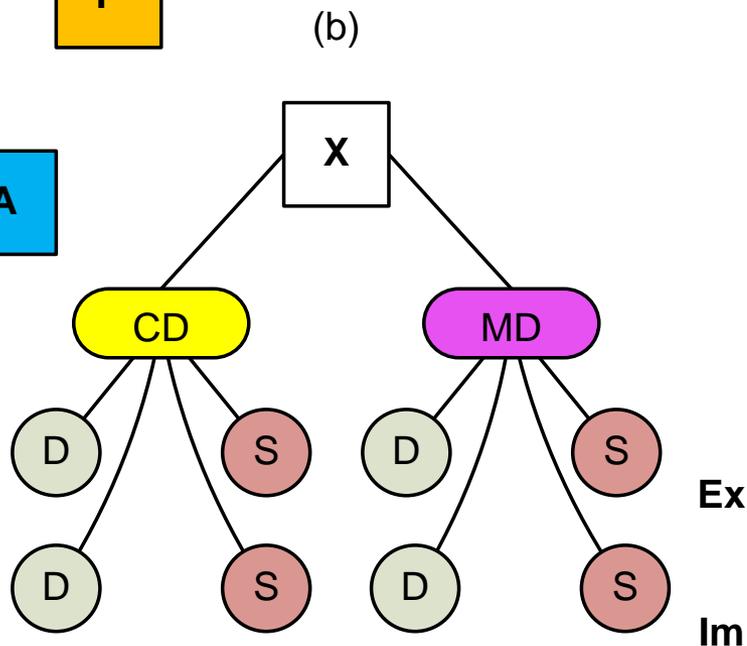
□ Графовые модели. Топология данных.

Графовые модели в анализе Социальных Медиа.

Основная причина появления графовых моделей при анализа соц. медиа – это отношения между объектами-контейнерами (см. лекцию I).

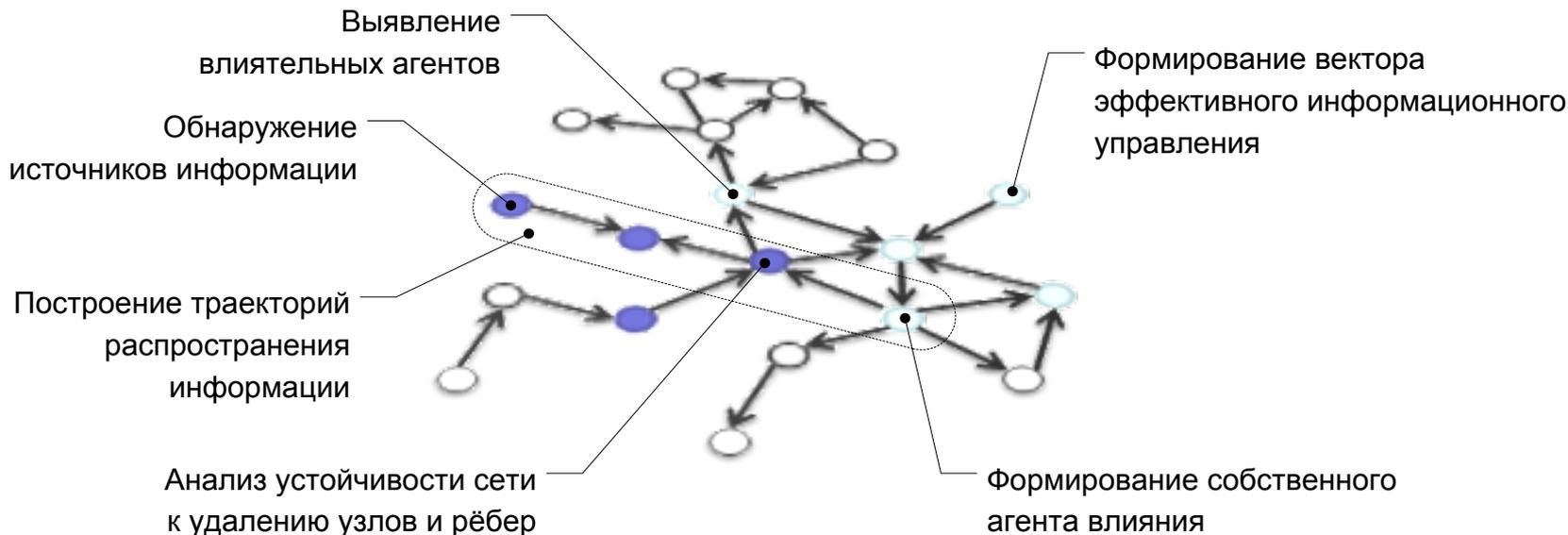


E – (enviroment) – внешняя среда;
P – (platform) – платформа отдельно взятого ресурса;
A – (account) – аккаунт пользователя;
M – (message) – сообщение пользователя;
B – (block) – логически целостная часть сообщения.

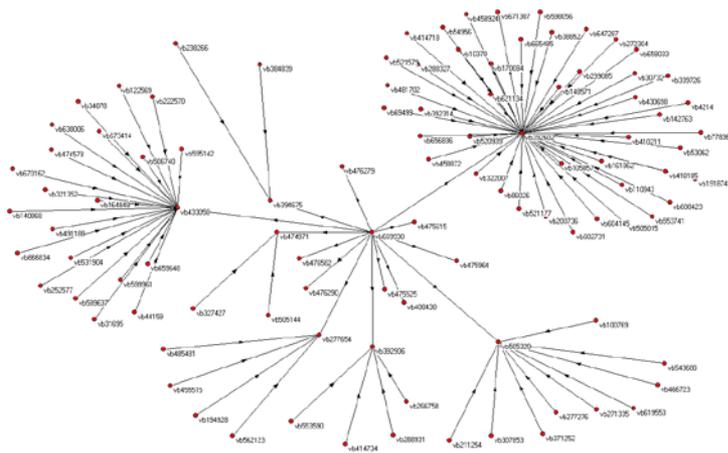


Графовые модели. Топология данных.

Графовые модели в анализе Социальных Медиа.



Всего около 25 главных функционально-ориентированных задач



 Facebook

 LinkedIn

 Twitter

 Newsvine

 Habr



□ Графовые модели. Топология данных.

Графовые модели в анализе Социальных Медиа.

Основные характеристики графов:

- Степени вершин, длины путей, диаметр графа, ...
- Центральность
 - по степени (degree centrality) – смысл этой меры основан на допущении, что тот, кто обладает большим количеством связей (отношений) с другими, занимает центральное положение в локальной общности.
 - по близости (closeness-centrality) – выражает, насколько близко узел расположен к остальным узлам сети или насколько легко достичь определенного узла в сети.
 - по посредничеству (betweenness-centrality) – выражает, сколько кратчайших путей между всеми узлами сети проходит через конкретный узел; характеризует, насколько важную роль данный узел играет на пути между другими узлами.
- Клики, мосты, связность, ...

□ Графовые модели. Топология данных.

Графовые модели в анализе Социальных Медиа.

```
> nodes <- read.csv("Dataset1-Media-Example-NODES.csv",
  header = T, as.is = T)
> links <- read.csv("Dataset1-Media-Example-EDGES.csv",
  header = T, as.is = T)
>
> links <- aggregate(links[, 3], links[, -3], sum)
> links <- links[order(links$from, links$to), ]
> colnames(links)[4] <- "weight"
> rownames(links) <- NULL
>
> head(nodes)
  id          media media.type type.label audience.size
1 s01          NY Times          1 Newspaper          20
2 s02 Washington Post          1 Newspaper          25
3 s03 Wall Street Journal          1 Newspaper          30
4 s04          USA Today          1 Newspaper          32
5 s05          LA Times          1 Newspaper          20
6 s06 New York Post          1 Newspaper          50
>
```

□ Графовые модели. Топология данных.

Графовые модели в анализе Социальных Медиа.

```
> head(links)
  from to      type weight
1  s01 s02 hyperlink    22
2  s01 s03 hyperlink    22
3  s01 s04 hyperlink    21
4  s01 s15  mention    20
5  s02 s01 hyperlink    23
6  s02 s03 hyperlink    21
>
> net <- graph.data.frame(links, nodes, directed = T)
> net <- simplify(net, remove.multiple = F, remove.loops = T)
>
```

□ Графовые модели. Топология данных.

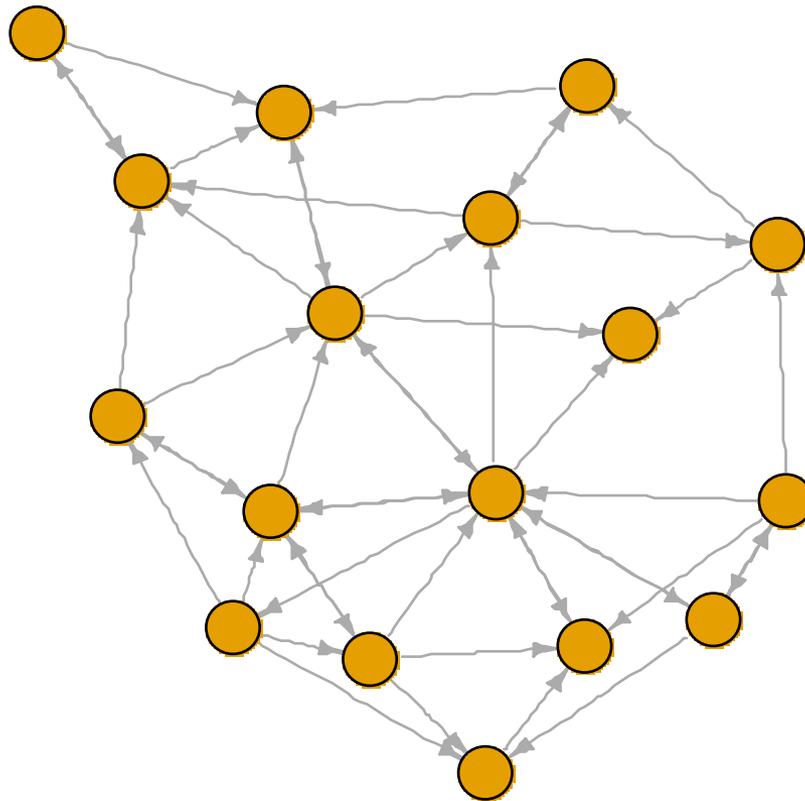
Графовые модели в анализе Социальных Медиа.

```
> net
IGRAPH DNW- 17 48 --
+ attr: name (v/c), media (v/c), media.type (v/n), type.label
  (v/c),
| audience.size (v/n), type (e/c), weight (e/n)
+ edges (vertex names):
[1] s01->s02 s01->s03 s01->s04 s01->s15 s02->s01 s02->s03 s02-
  >s09 s02->s10
[9] s03->s01 s03->s04 s03->s05 s03->s08 s03->s10 s03->s11 s03-
  >s12 s04->s03
[17] s04->s06 s04->s11 s04->s12 s04->s17 s05->s01 s05->s02 s05-
  >s09 s05->s15
[25] s06->s16 s06->s17 s07->s03 s07->s08 s07->s10 s07->s14 s08-
  >s03 s08->s07
[33] s08->s09 s09->s10 s10->s03 s12->s06 s12->s13 s12->s14 s13-
  >s12 s13->s17
[41] s14->s11 s14->s13 s15->s01 s15->s04 s15->s06 s16->s06 s16-
  >s17 s17->s04
>
```

□ Графовые модели. Топология данных.

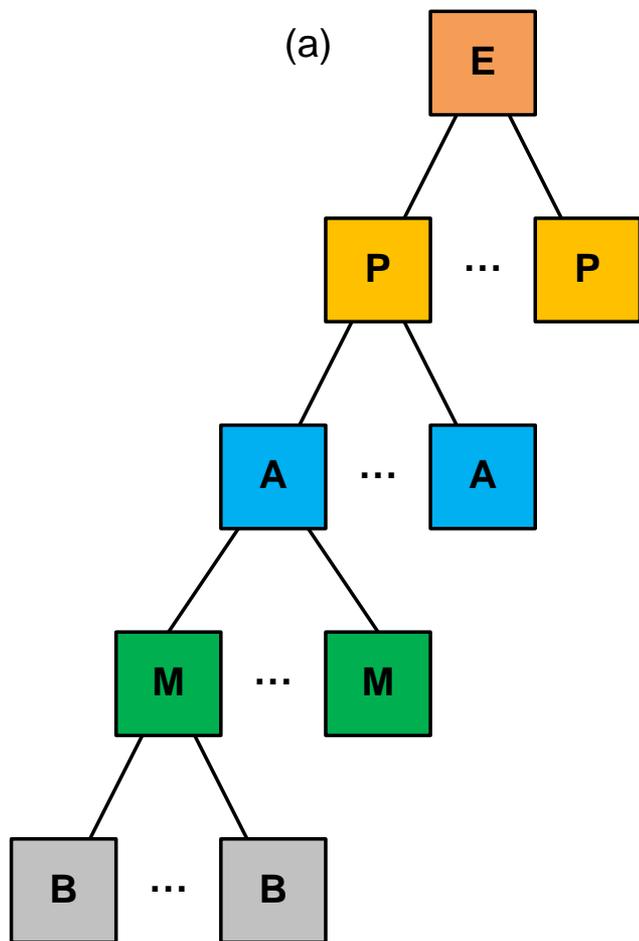
Графовые модели в анализе Социальных Медиа.

```
> par(mar = c(1, 1, 1, 1), mfp = c(5, 1, 0))  
> plot(net, edge.arrow.size = 0.4, vertex.label = NA)  
>
```



□ Вторичная формализация данных. Поиск информативных признаков.

Первичная структура данных.



E – (enviroment) – внешняя среда;

P – (platform) – платформа отдельно взятого ресурса;

A – (account) – аккаунт пользователя;

M – (message) – сообщение пользователя;

B – (block) – логически целостная часть сообщения.

Net: $n_i \in \mathbf{N} \subset \mathbb{N}$

Account: $g_j \in \mathbf{G} \mid n_i \subset \mathbb{N}$

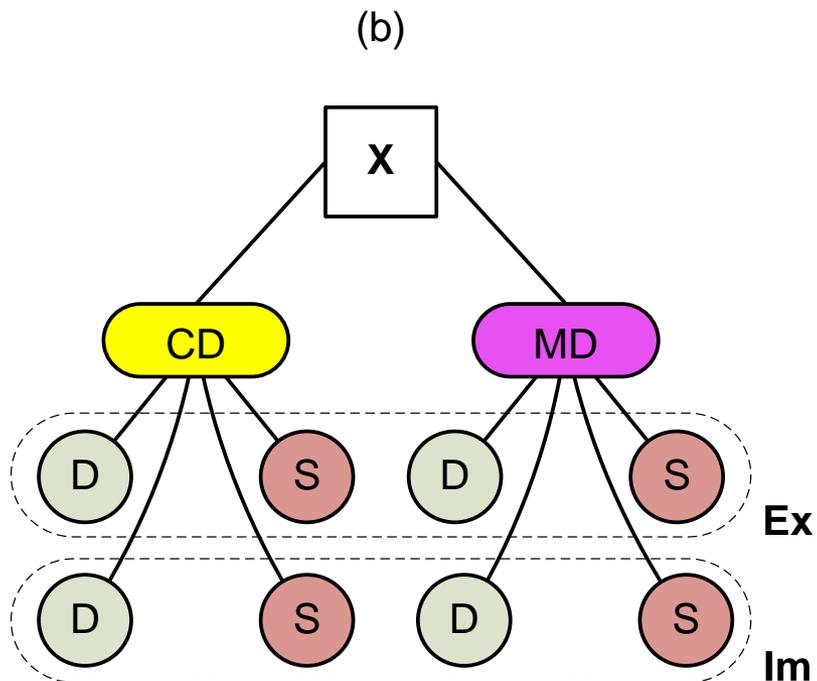
Message: $s_k \in \mathbf{S} \mid (g_j \mid n_i) \subset \mathbb{N}$

Block: $c_l \in \mathbf{C} \mid (s_k \mid (g_j \mid n_i)) \subset \mathbb{N}$

t_m^a – время порождения объекта, *важнейший элемент* метаданных.

□ Вторичная формализация данных. Поиск информативных признаков.

Вторичная структура данных.



Порождение вторичной структуры данных:

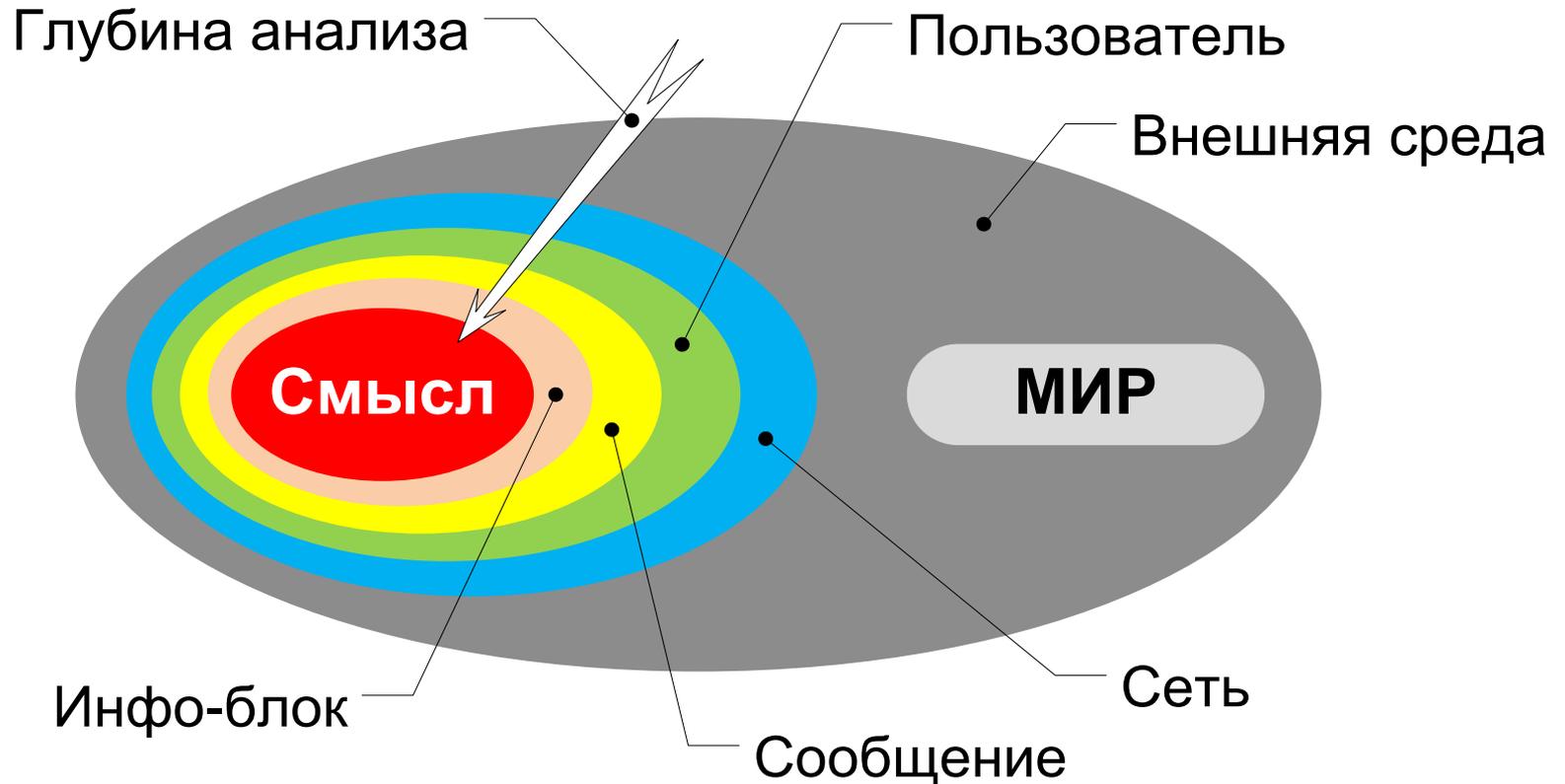
- Обработка неявных отношений
- Обработка текста
- Обработка картинок
- Обработка видео
- Обработка звука

Вторичная структура данных существенно усложняет обработку:

- Задачи компьютерной лингвистики, машинного зрения, распознавания голоса
- Размерность вектора и объём данных

□ Машинное обучение. Структурные модели. Вскрытие смысла.

Выделение сведений и Вскрытие смысла.



□ Машинное обучение. Структурные модели. Вскрытие смысла.

Машинное обучение по прецедентам.

Два основных подхода:

- Обучение с учителем (требует маркированной, размеченной выборки)
- Обучение без учителя (требует выборки существенного размера)

Основные решаемые задачи:

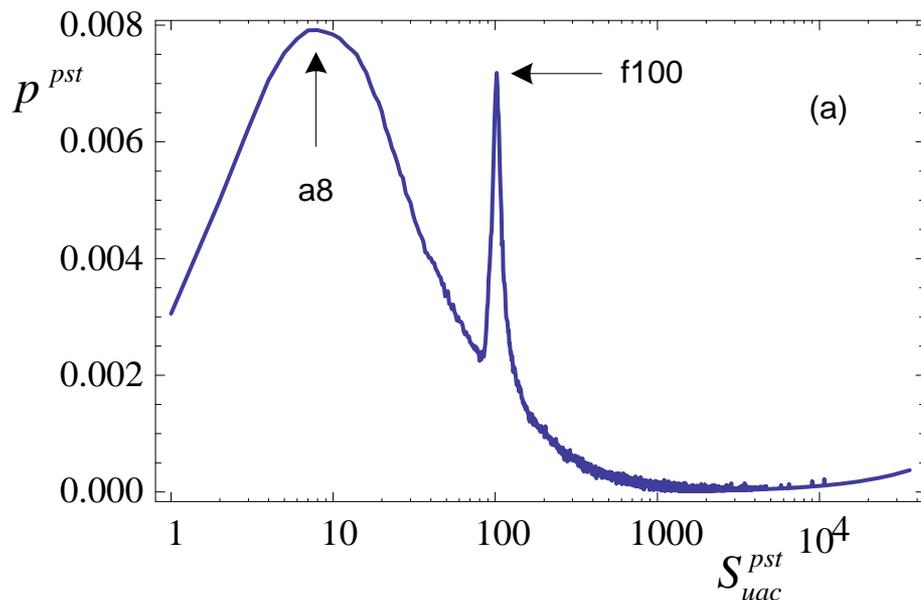
- Классификация
- Кластеризация
- Регрессия
- Понижение размерности данных
- Восстановление функции ПРВ по набору данных

Подробнее: <http://www.machinelearning.ru/wiki/index.php>

□ Когнитивная визуализация. Бизнес-инфографика.

Когнитивная визуализация.

Когнитивная визуализация – это визуализация, которая выполняет иллюстративно-информативную функцию и способствует естественно-интеллектуальному процессу получения новых знаний.



(!) Важное ограничение: **человеческий фактор**

Основные решаемые задачи:

- Выявление скрытых закономерностей
- Уточнение априорных моделей систем и процессов
- Выделение информативных признаков

Важные моменты:

- Формат визуализации
- Релевантные переменные

□ Задачи для самостоятельного решения

Теоретического плана (отчёт в электронном виде):

1. Возможно ли задачу регрессии представить в виде задачи классификации? Привести доводы и примеры обосновывающие ту или иную точку зрения.

Практического плана:

1. Изучить основы программирования на языке R.
2. Изучить пакет `igraph` R. Вывести метки вершин на изображение графа со слайда № 42.
3. Проанализировать состав пакетов R, доступных в официальном репозитории. Какие из пакетов могут быть полезны для анализа социальных медиа? В каких случаях и аспектах?