

# Оценка качества поисковых систем

## Лекция 9

БГУ ФПИИ, 2018

# План

Аспекты качества поисковых систем

Оффлайн метрики качества

Онлайн тестирование

# План

## Аспекты качества поисковых систем

### Оффлайн метрики качества

### Онлайн тестирование

## Аспекты качества поисковых систем

- ▶ Эффективность — как хорошо система находит результаты.
- ▶ Производительность — как быстро система находит результаты.

# Производительность

- ▶ RPS
- ▶ Время построения индекса
- ▶ Время ответа
- ▶ Размер индекса

# Эффективность

- ▶ Точность
- ▶ Полнота
- ▶ Ранжирование

# Метрики

- ▶ Оффлайн ( $F$ -мера, NDCG, ERR, AUC, ...)
- ▶ Онлайн (Switch, кликовые счетчики)

# План

Аспекты качества поисковых систем

Оффлайн метрики качества

Онлайн тестирование



## Задача предсказания релевантности

Обучающая выборка:

$$X = (q_i, d_i, r_i)_{i=1}^n$$

$F(q, d) \in \mathcal{F}$  – ранжирующая функция.

$$F(q, d) = \operatorname{argmax}_{F \in \mathcal{F}} \sum_{i=1}^n Q(F(q_i, d_i), r_i)$$

- ▶ задача классификации:  $r_i$  – дискретная метка
- ▶ задача регрессии:  $r_i$  – действительное число

## ROC-кривая

Рассмотрим задачу предсказания бинарной релевантности:  $r_i \in \{0, 1\}$ . Пусть  $F(q, d)$  возвращает некоторую непрерывную монотонную величину, по значению которой определяется принадлежность объекта к классу 1

$$F(q, d) > w \Rightarrow \hat{r}_i = 1$$

## ROC-кривая

Это график зависимости доли верных положительных классификаций от доли ложных положительных классификаций при варьировании порога  $w$ .

# AUC

- 1:  $\ell_- := \sum_{i=1}^{\ell} [y_i = -1]$  — число объектов класса  $-1$ ;  
 $\ell_+ := \sum_{i=1}^{\ell} [y_i = +1]$  — число объектов класса  $+1$ ;
- 2: упорядочить выборку  $X^\ell$  по убыванию значений  $f(x_i)$ ;
- 3: поставить первую точку в начало координат:  
 $(\text{FPR}_0, \text{TPR}_0) := (0, 0)$ ;  $\text{AUC} := 0$ ;
- 4: **для**  $i := 1, \dots, \ell$
- 5:   **если**  $y_i = -1$  **то**
- 6:     сместиться на один шаг вправо:  
       $\text{FPR}_i := \text{FPR}_{i-1} + \frac{1}{\ell_-}$ ;  $\text{TPR}_i := \text{TPR}_{i-1}$ ;  
       $\text{AUC} := \text{AUC} + \frac{1}{\ell_-} \text{TPR}_i$ ;
- 7:   **иначе**
- 8:     сместиться на один шаг вверх:  
       $\text{FPR}_i := \text{FPR}_{i-1}$ ;  $\text{TPR}_i := \text{TPR}_{i-1} + \frac{1}{\ell_+}$ ;

## Задача оптимального ранжирования

$$X = (q_i, \langle d_{i1}, r_{i1} \rangle, \dots, \langle d_{ik}, r_{ik} \rangle)_{i=1}^n$$

$F(q, D) \in \mathcal{F}$  – ранжирующая функция.

$$F(q, D) = \operatorname{argmax}_{F \in \mathcal{F}} \sum_{i=1}^n Q(F(q_i, D), \langle d_{i1}, r_{i1} \rangle, \dots, \langle d_{ik}, r_{ik} \rangle)$$

## Точность и полнота

### Точность

$$P = \frac{|D_{rel} \cap D_{retr}|}{|D_{retr}|}$$

### Полнота

$$R = \frac{|D_{rel} \cap D_{retr}|}{|D_{rel}|}$$

### F-мера

$$F = \frac{(1 + \beta^2)PR}{\beta^2 P + R}$$

## MAP

$$AP@k(q, d_1, \dots, d_j) = \frac{\sum_{j=1}^k P(k) \cdot rel(q, d_j)}{|D_{rel}|}$$

$$MAP@k = \frac{\sum_{i=1}^n AP@k(q_i)}{|Q|}$$

## NDCG

$$DCG_p = rel_1 + \sum_{i=2}^p \frac{rel_i}{\log_2 i}$$

$$DCG_p = \sum_{i=2}^p \frac{2^{rel_i} - 1}{\log_2(1 + i)}$$

$$NDCG@k = \frac{DCG@k}{IDCG@k}$$

# План

Аспекты качества поисковых систем

Оффлайн метрики качества

Онлайн тестирование



## A/B тест

- ▶ Пользователям из экспериментальной части потока (1-10%) показывается вариант с изменением, остальной части потока – предыдущая версия сервиса.
- ▶ По каждому из вариантов считаются статистики.
- ▶ При достаточном количестве наблюдений оценивается статистическая значимость изменения.