

Прогноз рейтингов рекламных блоков

Постановка задачи

Задача

Дано:

- Исторические рейтинги минутных и получасовых слотов для данного телеканала и данной целевой аудитории
- Историческая сетка вещания данного телеканала
- Описания (разнородные) объектов показа из исторической сетки

Найти:

- Рейтинг объекта показа на момент окончания показа (до начала рекламного блока) для данного канала, данной целевой аудитории

NB, далее:

все высказывания подразумевают, что

задана целевая аудитория и

задан телеканал

Варианты решения

Предлагаются два алгоритма прогноза:

- по прецедентам

«показывают всегда одно и то же»

- **ДЕКОМПОЗИЦИЯ → КОМПОЗИЦИЯ**

(время важнее наполнения)

«важно не что показывают, а когда показывают»

по прецедентам

Дана таблица «объект-описание»

г ^{ист}	объект	тип	дата	время	погода	...
------------------	--------	-----	------	-------	--------	-----

В список описаний входят как описание типа объекта, например, «спорт», «новости», так и время показа, например, «зима», «воскресенье», «новый год», «21:00».

Каждый признак метризован, например:

описание «время» («прайм», «полдень») = 0;

описание «тип» («спорт», «спорт») = 1.

Сходство объектов показа

Сходство объектов показа задано как **логическое правило**,

- 1) определенное на расстояниях по набору метризованных показателей, либо
- 2) определенное непосредственно на значениях показателей

Прогноз строится по историческим рейтингам сходных объектов

например, средняя (или взвешенная) сумма рейтингов

Допущение: объекты показа «не устаревают»

Поиск логического правила

Правило отыскивается и проверяется по исторической выборке, согласно одному из критериев качества, например:

рейтинг сходных объектов должен отличаться не более чем на ϵ

NB: поиск выполняется на подмножестве всех правил, с ограничениями, накладываемыми экспертом

Качество прогноза

обозначим $\Delta r = | r \text{ измеренный} - r \text{ прогнозируемый} |$

$$\sum \Delta r^2(t) \quad t=1, \dots, T$$

$$\max \Delta r^2(t)$$

$$\sum \Delta r^2(t) m(t)$$

$$\max \Delta r^2(t) m(t)$$

$1, \dots, T$ – отрезок времени

$m(t)$ – стоимость рекламного времени, нормированная относительно числа показов

Декомпозиция

Рейтинг слота зависит от:

- Качества объекта вещания
- Объема аудитории телеканала
- Вероятность того, что зритель будет смотреть ТВ в данное время

$$r = r(q_k, p_t)$$

q_k – качество объекта вещания с ID = k

p_t – вероятность работы телевизоров (аудитории) в момент t

ДЕКОМПОЗИЦИЯ → КОМПОЗИЦИЯ

Основная идея алгоритма: вычислить отдельно функции q_k, p_t и построить формулу прогноза рейтинга

Способ 1.

Задано приближение q_k, p_t ;

вычисляем $q'_k = r_{tk} / p_t$; $p'_k = r_{tk} / q_t$

корректируем q_k, p_t с учетом q'_k, p'_k

итеративно повторяем до сходимости (если есть)

ДЕКОМПОЗИЦИЯ → КОМПОЗИЦИЯ

Способ 2.

Строим универсальную модель, (например, нейросетевую), от двух переменных, k и t .

Замечание: оба способа предполагают, что каждая телепередача имеет индекс k . Этот индекс есть индекс класса эквивалентности телепередач. До начала работы алгоритма требуется каждой передаче поставить в соответствие индекс, пользуясь описанием телепередачи.

Вопросы

1. Каково типичное число представителей разных целевых аудиторий?
2. Вероятность того, что зритель будет смотреть ТВ в данное время: можно ли эту величину оценить по стоимости рекламного времени?
3. Рейтинг канала – число смотрящих канал в момент времени t или усредненное число смотрящих?
4. Рейтинг канала в выходные зависит от погоды? Какие дополнительные данные можно привлечь?