

Metric definition Overview

1. invariant checking 不变量检测

指标在实验组和对照组中都不会更改

e.g. 群数，分布。

2. evaluation 评估 revenue

1. high level business metric: 收益, market share, user.

2. detail : user experience, 停留时间

how to define?

1. 为一个指标想出高级概念, 大家通过这句话能够理解

这个指标 - e.g. active user, click probability

2. 确定各种细节。

e.g. how to define active?

3. summarized them as single metric: e.g. avg

如果有多个指标，可以创建一个复合指标。composite metric
称为 objective function or DEC 综合评估标准。

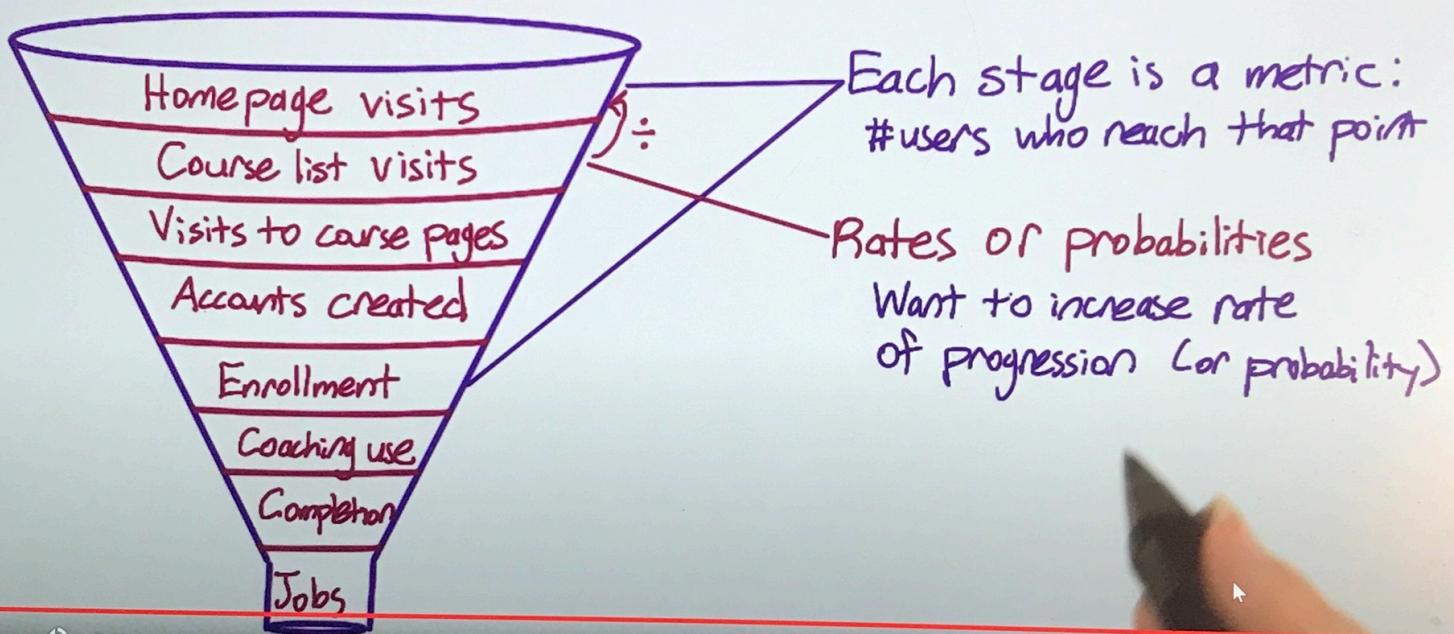
△ 考虑指标是否普遍适用。

Expanding on the funnel

- Homepage visits
- Exploring the site
 - # users who view course list
 - # users who view course details
- Create an account
 - # users who enroll in a course
 - # users who finish Lesson 1, Lesson 2, etc
 - # users who sign up for coaching at various levels
- Completing a course
 - # users who enroll in a second class
 - # users who get jobs

Expanding on the funnel

Different platforms?



Difficult metrics:

- Don't have access to data
- Takes too long.

↳ ~~B2B~~ R&M technique:

survey, retrospective analyses, focus group

↳ 用于 brainstorming new metrics & validating possible metrics.

Company collects data such as:

- ① market share, vertical (how many people are interested in the travel industry and even for specific website).
- ② company that run surveys of user
- ③ academic research:

company their data.

- ① uses all of our existing data:
- ② When you want to gather new data,
→ user experience research, surveys, and focus groups.

→ logging data or data capture from site.

Step 1: retrospective / observational analysis:

where you look at how a metric you're interested in

or measurements you take from your site, change in response to changes you've made in the past, experiments you've run, big spikes in your business.

目的：get a baseline & help you develop theories.

同时可以跟 surveys / user experience research.

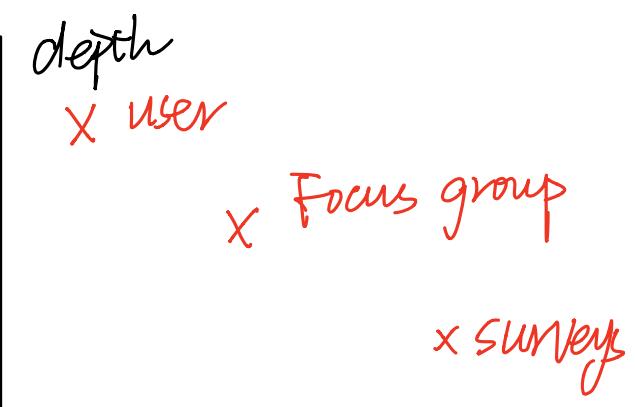
Gathering additional data

• User experience research (USER)

+ Good for brainstorming

+ can use special equipment

- hard to validate results.



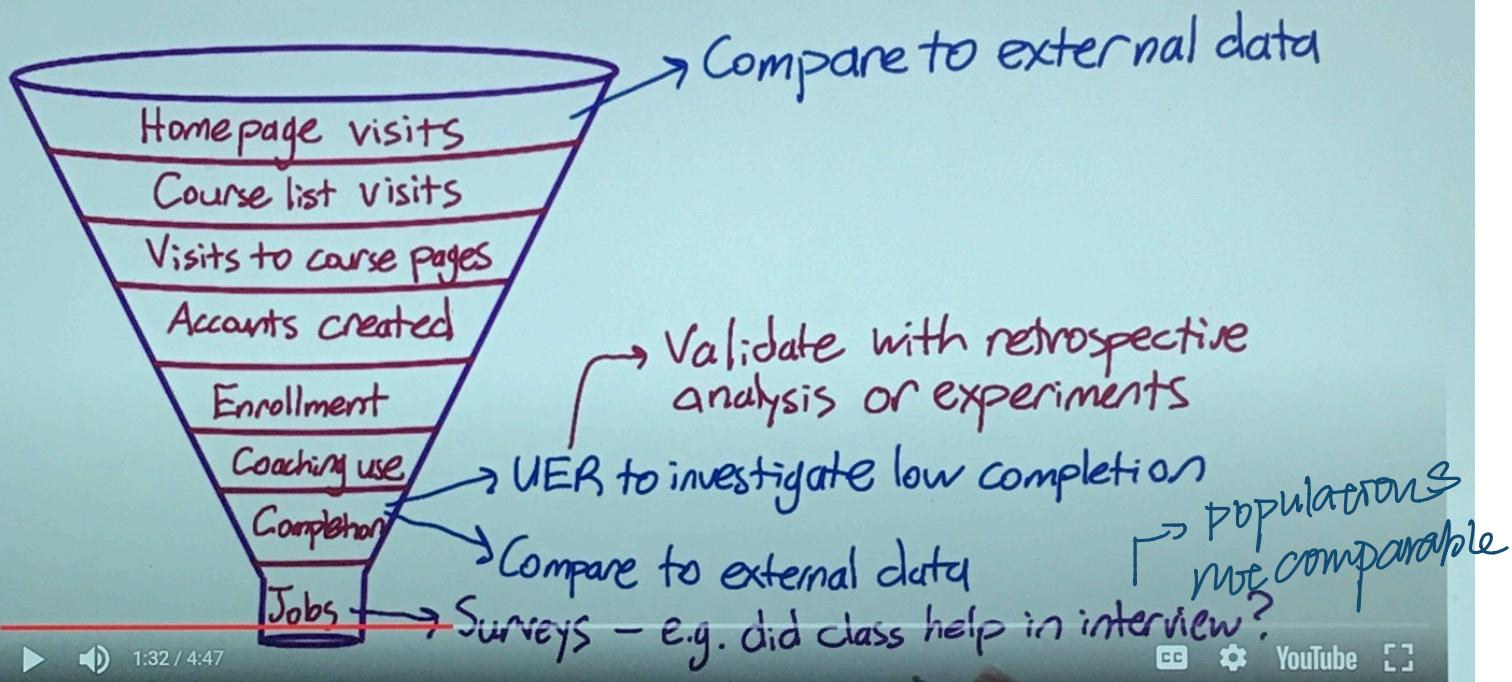
• Focus groups

- + Get feedback on hypotheticals # participants
- run the risk of group think.

Survey

- + useful for metric you cannot directly measure.
- cannot directly compare to other results

Applying other techniques



Applying other techniques to different metrics:

- Rate of returning for 2nd courses.
 - Survey → proxy
 - Average happiness of shoppers.
 - Survey
 - UER
 - Probability of finding information via research
 - External data
 - UER
 - Human evaluation.
- ↓
- possible proxies:
- time spent
 - clicks on results
 - follow-up queries

Metric definition and data capture.

怎样 define metric

怎样 summarize metric (e.g. mean, median,).

► 是否理解数据和指标中的度量.

Defining a metric:

High-level metric: click-through-probability = $\frac{\# \text{ user who click}}{\# \text{ user who visit}}$

Def #1: For each <time interval>, $\frac{\# \text{ cookies that click}}{\# \text{ cookies}}$



clicks no click click no click

$$\text{per minute: } \frac{2}{3} \quad \text{per hour: } \frac{1}{2} \quad \text{per day: } 1$$

Def #2: $\frac{\# \text{ pageviews w/ click within } <\text{time interval}>}{\# \text{ pageviews}}$



pageview pageview click. $\text{def}_1 = \frac{1}{1} = 1$

$$\text{def}_2 = \frac{1}{2}$$

Def #3 : $\frac{\# \text{clicks}}{\# \text{pageviews}}$ (click-through-rate)

which metrics have which problems?

check the metrics that would remain the same even if the problem occurred:

	1. cookie prob	2. pageview prob	3. rate
Double click.	✓	✓	
Back button caches page	✓		
click-tracking bug			

. 1min . ^{30sec}

time interval \rightarrow 5 min

情况1: pageview click click. both probability = 1
rate = 2.

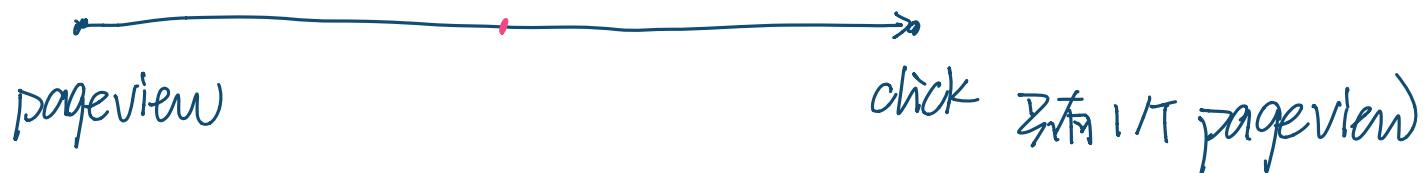
情况 2：

在第一条时间线上：

有 $2/T$ pageview



第二条时间线上，the browser caches the page so pageview is missing



情况 3：



Filtering and segmenting \rightarrow building intuition

筛选流量 { 外因： { 竞争对手 } 垃圾邮件 / 欺诈信息. \hookrightarrow what changes are expected

内因： e.g. 侧边栏 change 只影响部分 traffic

Bias: don't bias your data

通常做法: filtering out very long/wired session of user's behavior.
在做之前,需要检查并确保不是你的网站. metric (指标), 或
logging (日志) that causing these session come up.

怎么做?

1. Create a base value for your metric.
2. Slice the data. — you're computing your metric on a bunch of disjoint sets. (e.g. by country, language, platform)
3. Complete your metric on all of these slices.
4. When looking at the filter, what you want to see is whether or not you're moving traffic disproportionately (不成比例) from one of these places or not.

⇒ 如果成比例 → good.

不成比例 → maybe biasing your results further.

week over week
 day over day
 slicing down to individual user group

} identify spam or fraud

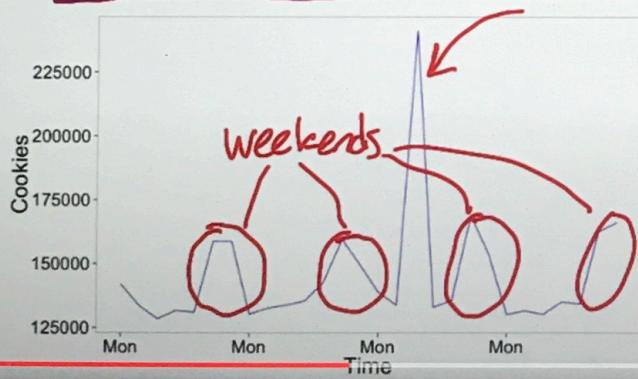
Segmenting and filtering data

Suspect click tracking issue : clicks counted twice on mobile.

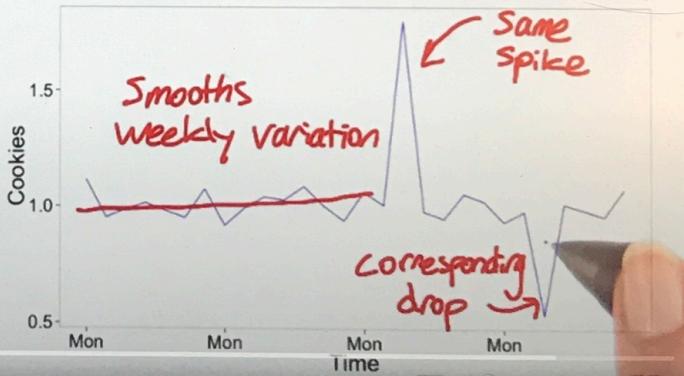
Segmenting and filtering data

Good for evaluating definitions and building intuition

Total active cookies over time



Cookies week-over-week



1. day-to-day. 周末访客数量更多.

2. 核实高峰是否是偶然情况:

week-over-week: 用每个数据点除以一周之前对应的数据点

3. week-to-week:

每周早晚变化 →

有一个 same spike, 说明高峰不是由于 weekly variation. →

Corresponding drop. because we divided this data point by the spike that occurred a week earlier.



说明 spike 也不是因为
yearly variation.

⇒ 解释方法:

looking at this metric of different segments of our population

to see if one segment is causing the spike.

$$\text{click-through-rate} = \frac{\text{click}}{\text{impression}}$$

$$\text{click-through probability} = \frac{\text{number of user who actually participate}}{\text{number of participants}}$$

Summary metrics.

↳ summarize all of these individual events into a single summary metric.

⇒ establish a few characteristics for your metric:

1. the sensitivity and robustness

want your metric to be sensitive enough to detect a change when you're testing your possible future options.

2. characterize is what the distribution of your metric looks like.

⇒ do a retrospective analysis and to compute a histogram.

histogram: on the x axis, you have all of the different values

for your metric. Y-axis is going to be the frequency.

e.g. X = load time

Y = how often individual event have that particular load time.

检查结果: What that shape look like.

e.g. normal shape \rightarrow mean / median.

one sided / lopsided \rightarrow 25th, 75th, 90th percentile.

① sums and counts ^{e.g.} # users who visited pages. percentiles

② all of the distributional metrics: means, medians, 25th, 75th, 90th

③ probabilities and rates \hookrightarrow e.g. mean age of users who completed a course.

④ ratios. — compute a whole range of different business models, and

various different things you care about, difficult to characterize.

probability has 0 or 1 outcome in each case.

rate has 0 or more.

④: any two numbers divided by each other.