

A/B Studies and Field Experiments

Intro

Seen so far:

- Introducing online evaluation
- Evaluating with usage logs

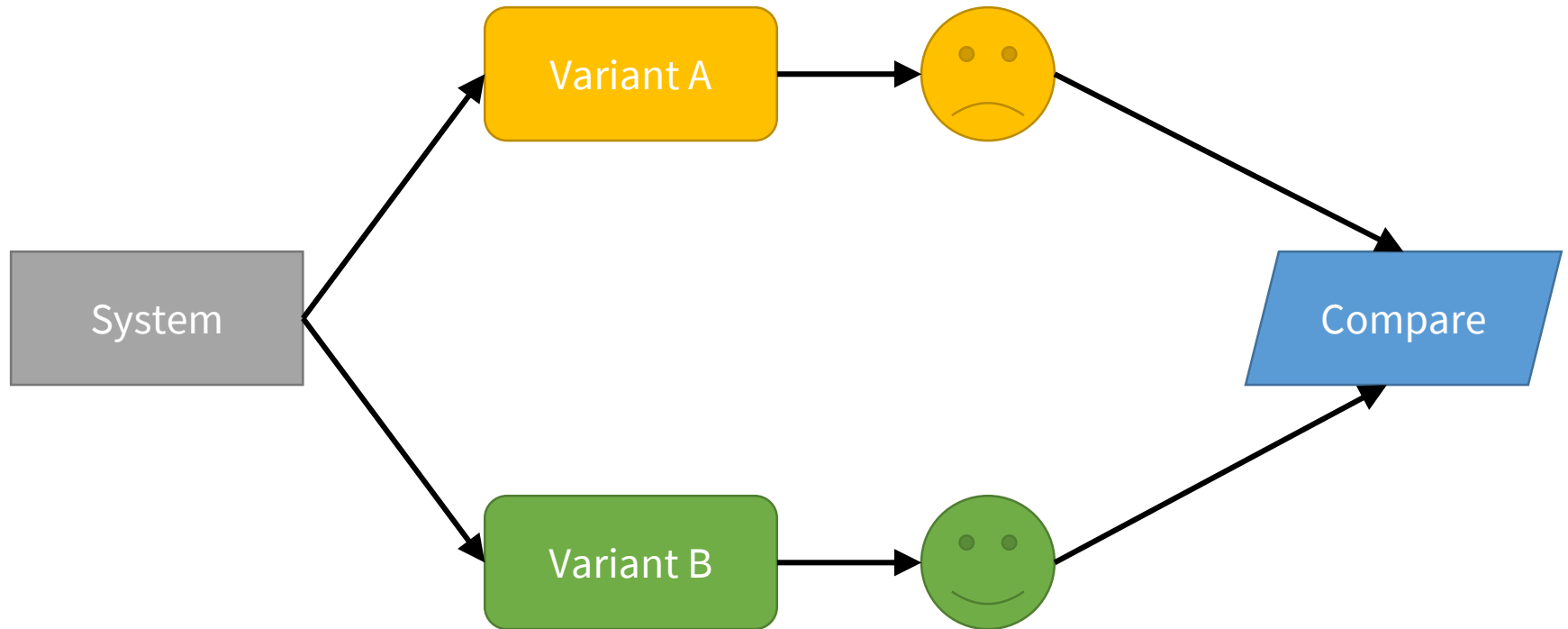
Now: designing and analyzing online experiments

Goal of Online Experiment

To see if a system change makes a positive improvement in user activity.

Not just applicable to recommenders.

Structure



Benefits

- Examine impact on actual user behavior
- Controlled experiment results in reliable knowledge
- Can scale out to multiple variants

Measuring Impact

Focus on a small number of **Key Performance Indicators** (KPIs)

- Videos watched
- Sales volume
- Conversion rate (recommendations that result in sales)

Selecting KPIs

Pick your KPIs carefully!

- Measure the **right thing**
- Don't incentivize **errant behavior**

Good KPIs measure **long-term customer value**

- Or a credible near-term proxy for it

Instrument Heavily

Do not *only* measure your KPIs

- Measure bounce rate (users leaving)
- Measure intermediate things
 - Page clicks/views aside from final outcomes

Why?

- Detect bad manipulations early
- Diagnose *why* experiment did or didn't work

Running Time

Don't just run a couple days & measure!



- Have 1-2 week burn-in period
 - Record data, but don't analyze for effect
 - Goal: flush out novelty effect
 - Important for anything new/visibly different!
 - Stop experiment if users react badly!
- Of course, targeting new users is different

Measurement Time

Measure for **multiple full weeks**

- Different days have different patterns (Tuesday afternoon isn't Saturday night)
- Weeks can be different
 - Be careful about special weeks

Analyzing Results

- Look for the manipulation's impact on measurable user behavior
- Analyze with robust statistics

Conclusion

- A/B testing lets you see impact of your recommender (or other change) on real users
- Powerful ability to measure recommender's impact on your business

See link in resources for more information!

A/B Studies and Field Experiments