

Instructions Homework Assignment 3

A/B Testing meets Exceptional Model Mining

December 05, 2018

A/B

Let's get Biblical! (Book of Daniel)

1:5 And the king appointed them a daily provision of the king's meat, and of the wine which he drank: so nourishing them three years, that at the end thereof they might stand before the king.

1:11 Then said Daniel to Melzar, whom the prince of the eunuchs had set over Daniel, Hananiah, Mishael, and Azariah, **1:12** Prove thy servants, I beseech thee, ten days; and let them give us pulse to eat, and water to drink. **1:13** Then let our countenances be looked upon before thee, and the countenance of the children that eat of the portion of the king's meat: and as thou seest, deal with thy servants. **1:14** So he consented to them in this matter, and proved them ten days. **1:15** And at the end of ten days their countenances appeared fairer and fatter in flesh than all the children which did eat the portion of the king's meat. **1:16** Thus Melzar took away the portion of their meat, and the wine that they should drink; and gave them pulse.

Core concepts in A/B Testing

We have:

- two variants of the same product;
- a pool of test subjects;
- a measure of success.

Core concepts in A/B Testing

We have:

- two variants of the same product;
- a pool of test subjects;
- a measure of success.

Procedure:

- 1 flip a coin for each test subject; assign corresponding variant;
- 2 aggregate success per variant;
- 3 variant with most success is kept, other is discarded.

Can we do more?

Final step of traditional A/B test: select winning version for the future, discard losing version.

Why discard, though? That seems unnecessarily wasteful!

Margin of victory is ignored. Subgroups that might be served better by the losing version are thrown under the bus.

Surely we can do more.

A / B

A & B

1. *Journal of Management Studies*, 1997, 34, 1, 1-14.



Surely, we can afford both A and B

Companies maintain several versions of websites all the time.

E.g., detect whether to give you the mobile or desktop version by browser settings.

Procedure of **A&B Testing**:

- 1 flip a coin for each test subject; assign corresponding variant;
- 2 aggregate success per variant;
- 3 variant with most success is kept as default;
- 4 further mine data to find coherent subgroups where alternative delivers more success;
- 5 present new visitors with either A or B depending on their encompassing subgroup.

So how do we mine the data further?

Country	OS	Height	Lang	viaGoogle	A/B	Click?
GR	Android	510	GR	0	A	No
NL	Android	615	EN-US	0	B	No
CH	iOS	955	EN-GB	0	A	No
IN	Android	232	EN-US	1	A	Yes
NG	Android	320	EN-US	0	B	No
IR	Android	460	FA	1	B	Yes

With **Exceptional Model Mining**!

Exceptional Model Mining

Supervised local pattern mining framework; multi-target.

Strives to find subsets of the dataset that:

- 1 are **interpretable**;
- 2 display **exceptional** target interaction.

Can be seen as generalization/extension of Subgroup Discovery.

Interpretable

Country	OS	Height	Lang	viaGoogle	A/B	Click?
GR	Android	510	GR	0	A	No
NL	Android	615	EN-US	0	B	No
CH	iOS	955	EN-GB	0	A	No
IN	Android	232	EN-US	1	A	Yes
NG	Android	320	EN-US	0	B	No
IR	Android	460	FA	1	B	Yes

Build **subgroups** from conditions on non-target columns.

Interpretability lives on the **left-hand side** of the dataset.

Interpretable

Country	OS	Height	Lang	viaGoogle	A/B	Click?
GR	Android	510	GR	0	A	No
NL	Android	615	EN-US	0	B	No
CH	iOS	955	EN-GB	0	A	No
IN	Android	232	EN-US	1	A	Yes
NG	Android	320	EN-US	0	B	No
IR	Android	460	FA	1	B	Yes

Country = GR

Interpretable

Country	OS	Height	Lang	viaGoogle	A/B	Click?
GR	Android	510	GR	0	A	No
NL	Android	615	EN-US	0	B	No
CH	iOS	955	EN-GB	0	A	No
IN	Android	232	EN-US	1	A	Yes
NG	Android	320	EN-US	0	B	No
IR	Android	460	FA	1	B	Yes

Country = GR

Height \leq 600

Interpretable

Country	OS	Height	Lang	viaGoogle	A/B	Click?
GR	Android	510	GR	0	A	No
NL	Android	615	EN-US	0	B	No
CH	iOS	955	EN-GB	0	A	No
IN	Android	232	EN-US	1	A	Yes
NG	Android	320	EN-US	0	B	No
IR	Android	460	FA	1	B	Yes

Country = GR

Height ≤ 600

Lang = EN-US \wedge viaGoogle = 0

Interpretable

Country	OS	Height	Lang	viaGoogle	A/B	Click?
GR	Android	510	GR	0	A	No
NL	Android	615	EN-US	0	B	No
CH	iOS	955	EN-GB	0	A	No
IN	Android	232	EN-US	1	A	Yes
NG	Android	320	EN-US	0	B	No
IR	Android	460	FA	1	B	Yes

Country = GR

Height \leq 600

Lang = EN-US \wedge viaGoogle = 0

Enumerate loads of subgroups, and move to the right-hand side.

Exceptional

Country	OS	Height	Lang	viaGoogle	A/B	Click?
GR	Android	510	GR	0	A	No
NL	Android	615	EN-US	0	B	No
CH	iOS	955	EN-GB	0	A	No
IN	Android	232	EN-US	1	A	Yes
NG	Android	320	EN-US	0	B	No
IR	Android	460	FA	1	B	Yes

Subgroups are **exceptional** if the target columns display unusual interaction, compared with overall behavior.

Exceptionality lives on the **right-hand side** of the dataset.

1

Country	OS	Height	Lang	viaGoogle	A/B	Click?
GR	Android	510	GR	0	A	No
NL	Android	615	EN-US	0	B	No
CH	iOS	955	EN-GB	0	A	No
IN	Android	232	EN-US	1	A	Yes
NG	Android	320	EN-US	0	B	No
IR	Android	460	FA	1	B	Yes

Country = GR

Height ≤ 600

Lang = EN-US \wedge viaGoogle = 0

Two categorical targets \Rightarrow association model class.

Association model class with Yule's Q

	No click	Click
A	n_1	n_2
B	n_3	n_4

$$Q = \frac{(n_1 \cdot n_4 - n_2 \cdot n_3)}{(n_1 \cdot n_4 + n_2 \cdot n_3)}$$

Positive Q: higher values on diagonal: more clicks when presented with B.

Negative Q: higher values off diagonal: more clicks when presented with A.

What are we going to do?

In this assignments you can earn points by:

- locating in the StudyPortals data the relevant attributes;
- preprocessing the multi-table dataset into single flat-table form suitable for EMM;
- implementing the beam search algorithm to mine for exceptional subgroups and associated models;
- implementing the specific association model class and Yule's Q quality measure to perform the EMM instance relevant for A&B testing;
- explore beyond the straightforward model class: what if we have more than two alternatives, what if we have a numeric instead of binary reward function, what if we need a significant difference instead of just any difference. . .

Who to contact in case of emergency?

If you have any questions about the dataset itself, contact Tara.

If you have any questions about the assignment:

- before Christmas, contact Hilde and/or Simon;
- in 2019, contact Wouter directly (via email).