Do You TURF?

An Optimization Approach

Agenda

▶ What is TURF analysis?

▶ How to…

- Gather TURF data

- Analyze TURF data in Sawtooth Software's MaxDiff Analyzer

▶ Reach vs. Overlap

TURF = Total Unduplicated Reach & Frequency

▸ Say WHAT???

▸ TURF is an optimization approach for finding a subset of items that "reach" the maximum number of respondents possible.

  ▪ "If we can only offer three flavors of ice cream, which three should we offer so that as many people as possible have at least one flavor that they like?"

What is TURF Analysis?

▶ Originated in media research to help maximize reach while minimizing media costs

▶ Ex. TV Show X has an audience of 1 million, TV Show Y has an audience of 2 million, it would be dangerous to assume if we advertise in X and Y we would get 3 million views

Applications

▶ What new item/flavor/service should I add to my product line to find new customers?

▶ Which mediums should I advertise in to reach the widest audience possible?

▶ How can I satisfy the most consumers with the fewest number of products/services?

Example

▸ Say we had 3 potential ice cream flavors (Vanilla, Chocolate, Mango) but we only want to produce or stock 2

▸ We might ask a rating question for each flavor (5 point scale?)

TURF Analysis

▶ Data set 1

  ▪ 75 respondents who rate Vanilla 5, Chocolate 4, Mango 1

  ▪ 25 respondents who rate Vanilla and Chocolate 1, Mango 5

▶ Data set 2

  ▪ 50 respondents who rate Vanilla 5

  ▪ 30 rate Chocolate 5

  ▪ 20 rate Mango 5

TURF Analysis

▸ Average scores for each data set

	Vanilla	Chocolate	Mango
Data Set 1 (overlap)	4	3.25	2
Data Set 2 (no overlap)	3	2.2	1.8

▸ Both data sets would support producing Vanilla and Chocolate

▸ TURF provides a different way of looking at the data to consider "reach" instead of average preference

TURF Analysis

▸ TURF considered reach and frequency in choosing optimal combinations

Resp.	Vanilla	Chocolate	Mango
1	1	1	0

▸ Respondent 1, who would buy both Vanilla and Chocolate, would produce the following TURF counts:

Offering	Reach	Frequency
Vanilla & Chocolate	1	2
Chocolate & Mango	1	1
Vanilla & Mango	1	1

TURF Analysis

▸ Adding respondent 2, who only would buy Mango, would create the following TURF counts:

Resp.	Vanilla	Chocolate	Mango
1	1	1	0
2	0	0	1

Offering	Reach	Frequency
Vanilla & Chocolate	1	2
Chocolate & Mango	2	2
Vanilla & Mango	2	2

▸ If we are aiming for reach, our optimal solution would include Mango since it would reach the greatest number of customers

So how do we get this TURF data?

▸ Multi-Select

   ▪ Ex. Anything given a 1 is considered "reached"

▸ Ranking

   ▪ Ex. Top 3 ranks are considered "reached" or top rank is considered "reached"

▸ Rating

   ▪ Ex. On a 5-point scale, top 2 box (4 or 5) are considered "reached"

▸ Count data*

   ▪ Ex. How many of each of the following would you consider purchasing? Where >0 is considered "reached"

   ▪ *Necessary if we want actual frequency reports

▸ MaxDiff Scores

Brief Reminder

MaxDiff (n). Maximum difference scaling, also known as Best Worst Scaling, is used for measuring importance, agreement, etc. within a list of items.

Example MaxDiff Question

▸ Considering only these four ice cream flavors, which is the Most Appealing and which is the Least Appealing?

(1 of 8)

Least Appealing		Most Appealing
○	Chocolate	○
○	Mint Chocolate Chip	●
●	Black Cherry	○
○	Mango	○

TURF Analysis With MaxDiff

▶ MaxDiff scores are continuous, rather than binary

 ▪ Option 1: "Descrete-ify" by using first-choice rule

▶ If their top item is in the bundle, they are reached

 ▪ Option 2: "Descrete-ify" by using a threshold

▶ If the score of an item exceeds threshold, that respondent is reached

 ▪ Option 3: Weighted by probability

First-Choice Criteria

▶ Example scores

	Vanilla	Chocolate	Mango
Resp 1	5.2	1.6	-4.7
Resp 2	-1.7	-5.3	4.8

▶ TURF output

Offering	Reach	Frequency
Vanilla & Chocolate	1	1
Chocolate & Mango	1	1
Vanilla & Mango	2	2

Threshold Criteria

▸ Example scores, threshold of 0

	Vanilla	Chocolate	Mango
Resp 1	5.2	1.6	-4.7
Resp 2	-1.7	-5.3	4.8

▸ TURF output

Offering	Reach	Frequency
Vanilla & Chocolate	1	2
Chocolate & Mango	2	2
Vanilla & Mango	2	2

Anchored MaxDiff – Dual Response Indirect Method

▶ Considering only these four ice cream flavors, which is the Most Appealing and which is the Least Appealing?

(1 of 8)

Least Appealing		Most Appealing
○	Cholate	○
○	Mint Chocolate Chip	●
●	Black Cherry	○
○	Mango	○

▶ Considering only the items above…

○ None of these are appealing to me

○ Some of these are appealing to me

● All of these are appealing to me

Anchored MaxDiff – Direct Binary Approach

▸ Asked after, or before the MaxDiff exercise as a separate question.

▸ Now, please tell us, which of the following flavors do you actually find appealing? (Select all that apply)

☐ Chocolate   ☐ Mango
■ Vanilla   ■ Superman
■ Strawberry   ☐ Blue Moon
☐ Mint Chocolate Chip   ■ Cookies & Cream
■ Black Cherry   ☐ None of the above

Weighted by Probability

‣ $P_i = e^U_i / (e^U_i + a - 1)$

‣ $P_{i+} P_{j=} e^U_i / (e^U_i + a - 1) + e^U_j / (e^U_j + a - 1)$

‣ $P_{ij} = (e^U_i + e^U_j) / (e^U_i + e^U_j + a - 1)$

Weighted by Probability Criteria

▶ Example Scores

	Vanilla	Chocolate	Mango
Resp 1	5.2	1.6	-4.7
Resp 2	-1.7	-5.3	4.8

▶ TURF output

Offering	Strength	Rank Order
Vanilla & Chocolate	0.536846358	3
Chocolate & Mango	0.842958873	2
Vanilla & Mango	0.985844141	1

MaxDiff Analyzer Demo

▶ https://www.maxdiffanalyzer.com

▶ New Improvements

- Updated look for the interface.

- You can now request a larger number of near-optimal portfolios in TURF (previously you always got 100; now it is adjustable up to 10000).

- We've changed the format for the downloaded output of TURF so that items in portfolios are separated by commas rather than by "&" symbols in the CSV file.

- Improved calculations for the weighted by probability method.

Report Examples
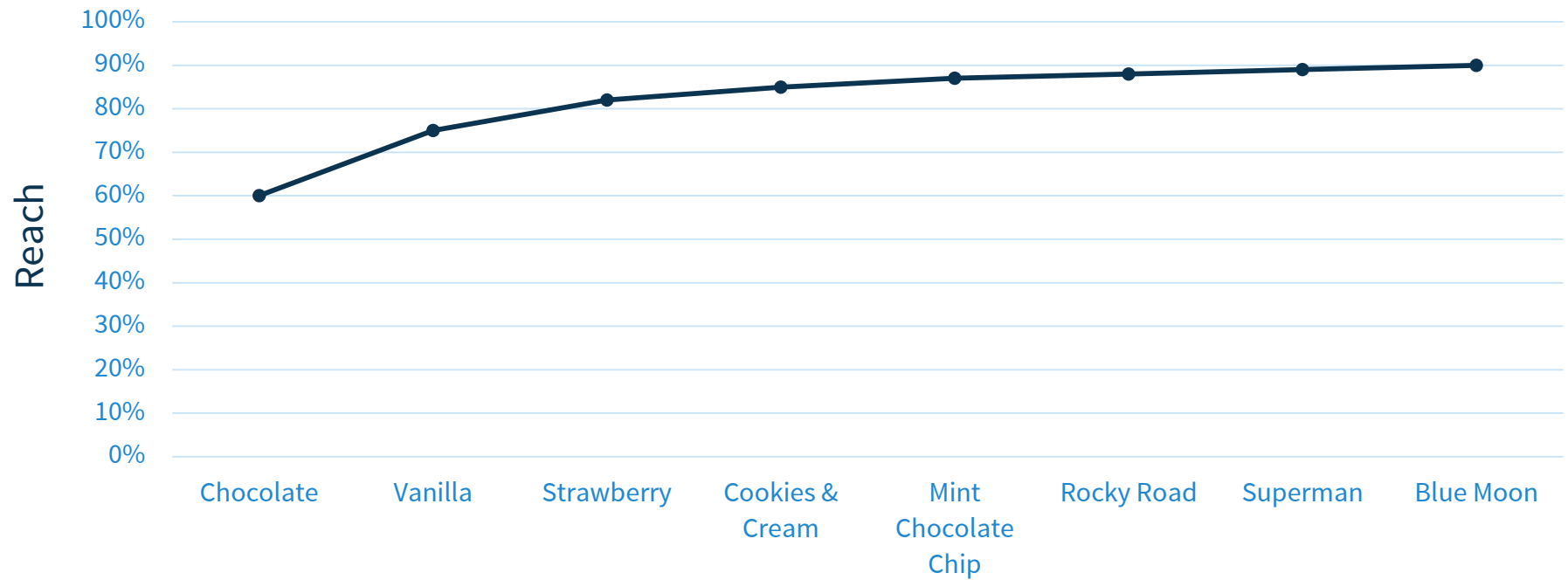
Limitations

▶ Consider adding weights to your data based on volume

- The frequency interpretation of TURF data is often misleading
- A person who is going to purchase 2x a week is much more valuable then someone who will purchase only once a month

▶ Our rules to create "reach" assume that a respondent is satisfied with one specific product and will not seek variety

- Be weary of just using TURF data in FMCG categories or other categories with variety seeking

▶ Consumer segments are seldom defined by one individual offering (ex. one credit card doesn't fit all, one flavor doesn't solve appeal to all occasions) – we need to make sure the entire product line is strong.

But what about overlap?

▶ While TURF reaches the largest amount of people, with the smallest amount of items, sometimes we want to make our item set the strongest.

- ▪ "If we can only offer three flavors of ice cream, which three should we offer so that as many people as possible have at least one flavor that they like?"

- ▪ But what if one of our strongest flavors is sold out? We don't want them to walk away from the shelf! We need a substitute or an item that "overlaps".

Association Rules

▶ Association rules are if/then statements that help uncover relationships between seemingly unrelated data.

- Ex. "If a customer buys a dozen eggs, he is 80% likely to also purchase milk."

▶ Use the aRules package in R:

- *Support* is an indication of how frequently the items appear in the database.

- *Confidence* indicates the number of times the if/then statements have been found to be true.
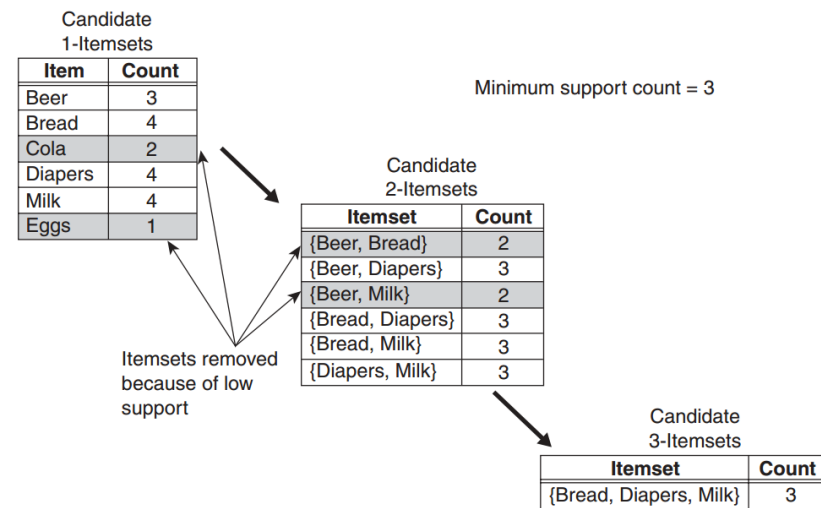

Figure 6.5. Illustration of frequent itemset generation using the *Apriori* algorithm.

Output Interpretation

▶ {milk, bread, butter} support=0.2

- Because this combination occurs in 20% of all transactions in the set

▶ {butter, bread} → {milk} confidence=100%

- For 100% of the transactions containing butter & bread, milk is also bought

$$\mathrm{conf}(X \Rightarrow Y) = \mathrm{supp}(X \cup Y)/\mathrm{supp}(X)$$

▶ {milk, bread} → {butter} lift=1.25

$$\mathrm{lift}(X \Rightarrow Y) = \frac{\mathrm{supp}(X \cup Y)}{\mathrm{supp}(X) \times \mathrm{supp}(Y)}$$

- If some rule had a lift of 1, it would imply that the probability of occurrence of the antecedent and that of the consequent are independent of each other. When two events are independent of each other, no rule can be drawn involving those two events.

- If the lift is > 1, that lets us know the degree to which those two occurrences are dependent on one another, and makes those rules potentially useful for predicting the consequent in future data sets.

Example in R

▶ http://www.rdatamining.com/examples/association-rules

```
  lhs                rhs                 support    confidence  lift
1 {Class=2nd,
   Age=Child}  => {Survived=Yes} 0.010904134  1.0000000 3.095640
2 {Class=1st,
   Sex=Female} => {Survived=Yes} 0.064061790  0.9724138 3.010243
3 {Class=2nd,
   Sex=Female} => {Survived=Yes} 0.042253521  0.8773585 2.715986
4 {Class=Crew,
   Sex=Female} => {Survived=Yes} 0.009086779  0.8695652 2.691861
5 {Class=2nd,
   Sex=Male,
   Age=Adult}  => {Survived=No}  0.069968196  0.9166667 1.354083
6 {Class=2nd,
   Sex=Male}   => {Survived=No}  0.069968196  0.8603352 1.270871
7 {Class=3rd,
   Sex=Male,
   Age=Adult}  => {Survived=No}  0.175829169  0.8376623 1.237379
8 {Class=3rd,
   Sex=Male}   => {Survived=No}  0.191731031  0.8274510 1.222295
```

QUESTIONS?

Brian McEwan
Director, Client Services
brian@sawtoothsoftware.com

Megan Peitz
Ingenuity Ambassador
megan@sawtoothsoftware.com

www.sawtoothsoftware.com
+1 801 477 4700
@sawtoothsoft