The apriori algorithm as an engine for computerized adaptive assessment

NIELS SMITS

Research Institute of Child Development and Education
University of Amsterdam, The Netherlands

PSYCHOCO, Dortmund
2020, February 28

Outline

Interest in alternative methods for adaptive testing

- Need for short self-report based assessments in health settings.

Interest in alternative methods for adaptive testing

- Need for short self-report based assessments in health settings.
- Assessment often aimed at classification or prediction.

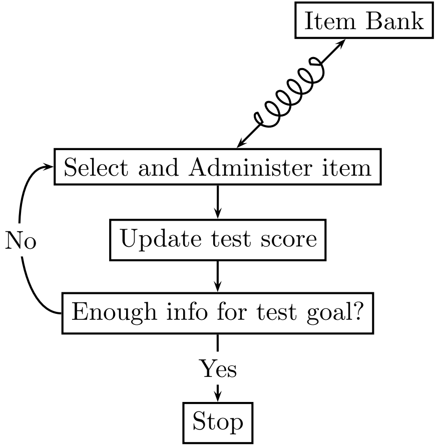
Interest in alternative methods for adaptive testing

- Need for short self-report based assessments in health settings.
- Assessment often aimed at classification or prediction.
- Such tests require specific construction approaches (Smits et al., 2018; Oosterveld et al., 2019).

Interest in alternative methods for adaptive testing

- Need for short self-report based assessments in health settings.
- Assessment often aimed at classification or prediction.
- Such tests require specific construction approaches (Smits et al., 2018; Oosterveld et al., 2019).
- Unfortunately, the standard approach under Item Response Theory is inappropriate.

Adaptive testing



Existing methods for classification and prediction

- Curtailment (a.k.a. ‘Countdown’, Butcher et al., 1985).

Existing methods for classification and prediction

- Curtailment (a.k.a. ‘Countdown’, Butcher et al., 1985).
- Stochastic Curtailment (Finkelman et al., 2012, 2013; Fokkema et al., 2014; Smits & Finkelman, 2015).

Existing methods for classification and prediction

- Curtailment (a.k.a. ‘Countdown’, Butcher et al., 1985).
- Stochastic Curtailment (Finkelman et al., 2012, 2013; Fokkema et al., 2014; Smits & Finkelman, 2015).
- But:
  - Early stopping, i.e., no dynamic item selection.
  - Focus on (cumulative) sum scores.

Requirements for classification and prediction

Method should:

- Provide sound approximation of cross tabulation of items.

Requirements for classification and prediction

Method should:

- Provide sound approximation of cross tabulation of items.
- Allow for predicting a criterion.

Requirements for classification and prediction

Method should:

- Provide sound approximation of cross tabulation of items.
- Allow for predicting a criterion.
- Allow for dynamic item selection.

Requirements for classification and prediction

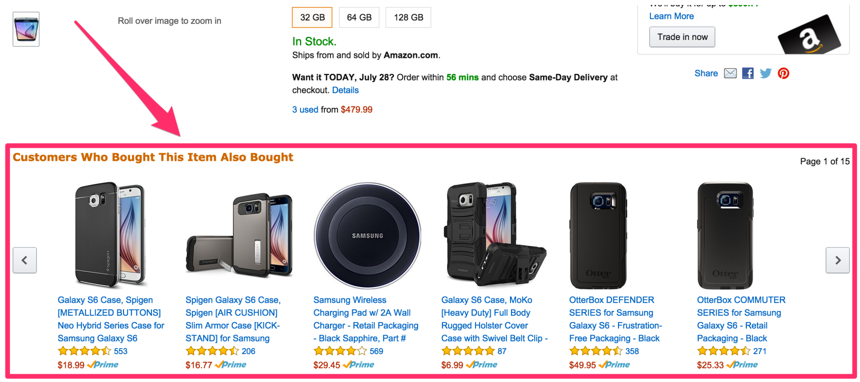
Method should:

- Provide sound approximation of cross tabulation of items.
- Allow for predicting a criterion.
- Allow for dynamic item selection.

Would a rule learning algorithm like `apriori` be useful?

Rule Learning: You already know this!

Rule Learning: You already know this!



Rule Learning: You already know this!

Rule Learning

- Association rules.

Rule Learning

- Association rules.
- Market Basket Analysis

Rule Learning

- Association rules.
- Market Basket Analysis
- What items are frequently bought together?

Rule Learning

- Association rules.
- Market Basket Analysis
- What items are frequently bought together?
- What symptoms frequently co-occur?

The Apriori Algorithm

Building blocks:

frequent set: $\mathcal{K} = A \cup B$.

The Apriori Algorithm

Building blocks:

frequent set: $\mathcal{K} = A \cup B$.
rule: $A \Rightarrow B$.

(A=antecedent, B=consequent.)

The Apriori Algorithm

Building blocks:

frequent set: $\mathcal{K} = A \cup B$.
rule: $A \Rightarrow B$.
support: $T(A \Rightarrow B)$.

(A=antecedent, B=consequent.)

The Apriori Algorithm

Building blocks:

frequent set: $\mathcal{K} = A \cup B$.

rule: $A \Rightarrow B$.

support: $T(A \Rightarrow B)$.

confidence: $C(A \Rightarrow B) = \dfrac{T(A \Rightarrow B)}{T(A)}$.

(A=antecedent, B=consequent.)

The Apriori Algorithm

Building blocks:

frequent set: $\mathcal{K} = A \cup B$.

rule: $A \Rightarrow B$.

support: $T(A \Rightarrow B)$.

confidence: $C(A \Rightarrow B) = \dfrac{T(A \Rightarrow B)}{T(A)}$.

lift: $L(A \Rightarrow B) = \dfrac{C(A \Rightarrow B)}{T(B)}$.

(A=antecedent, B=consequent.)

The Apriori Algorithm

Example:

$$\mathcal{K} = \{\texttt{sleeping},\ \texttt{eating},\ \texttt{concentration}\}.$$

The Apriori Algorithm

Example:

$$\mathcal{K} = \{\texttt{sleeping},\ \texttt{eating},\ \texttt{concentration}\}.$$

$$\{\texttt{sleeping},\ \texttt{eating}\} \Rightarrow \{\texttt{concentration}\}.$$

The Apriori Algorithm

Example:

$$\mathcal{K} = \{\texttt{sleeping, eating, concentration}\}.$$
$$\{\texttt{sleeping, eating}\} \Rightarrow \{\texttt{concentration}\}.$$
$$T(\{\texttt{sleeping, eating}\}) = 0.05.$$

The Apriori Algorithm

Example:

$$\mathcal{K} = \{\texttt{sleeping, eating, concentration}\}.$$

$$\{\texttt{sleeping, eating}\} \Rightarrow \{\texttt{concentration}\}.$$

$$T(\{\texttt{sleeping, eating}\}) = 0.05.$$

$$T(\{\texttt{concentration}\}) = 0.15.$$

The Apriori Algorithm

Example:

$$\mathcal{K} = \{\texttt{sleeping},\ \texttt{eating},\ \texttt{concentration}\}.$$
$$\{\texttt{sleeping},\ \texttt{eating}\} \Rightarrow \{\texttt{concentration}\}.$$
$$T(\{\texttt{sleeping},\ \texttt{eating}\}) = 0.05.$$
$$T(\{\texttt{concentration}\}) = 0.15.$$
$$T(\{\texttt{sleeping},\ \texttt{eating}\} \Rightarrow \{\texttt{concentration}\}) = 0.03.$$

The Apriori Algorithm

Example:

$$\mathcal{K} = \{\texttt{sleeping},\ \texttt{eating},\ \texttt{concentration}\}.$$
$$\{\texttt{sleeping},\ \texttt{eating}\} \Rightarrow \{\texttt{concentration}\}.$$
$$T(\{\texttt{sleeping},\ \texttt{eating}\}) = 0.05.$$
$$T(\{\texttt{concentration}\}) = 0.15.$$
$$T(\{\texttt{sleeping},\ \texttt{eating}\} \Rightarrow \{\texttt{concentration}\}) = 0.03.$$
$$C(\{\texttt{sleeping},\ \texttt{eating}\} \Rightarrow \{\texttt{concentration}\}) = 0.60.$$

The Apriori Algorithm

Example:

$$\mathcal{K} = \{\texttt{sleeping, eating, concentration}\}.$$
$$\{\texttt{sleeping, eating}\} \Rightarrow \{\texttt{concentration}\}.$$
$$T(\{\texttt{sleeping, eating}\}) = 0.05.$$
$$T(\{\texttt{concentration}\}) = 0.15.$$
$$T(\{\texttt{sleeping, eating}\} \Rightarrow \{\texttt{concentration}\}) = 0.03.$$
$$C(\{\texttt{sleeping, eating}\} \Rightarrow \{\texttt{concentration}\}) = 0.60.$$
$$L(\{\texttt{sleeping, eating}\} \Rightarrow \{\texttt{concentration}\}) = 4.00.$$

Building blocks

Requirements:

- Rule data base.
- Item selection.
- Test score.
- Stopping rule.

Rule data base

- Standard analysis focuses on *presence* of items.

Rule data base

- Standard analysis focuses on *presence* of items.
- For health assessment *absence* of symptoms also important.

Rule data base

- Standard analysis focuses on *presence* of items.
- For health assessment *absence* of symptoms also important.
- In calibration, both presence and absence included ('doubling').

Rule data base

- Standard analysis focuses on *presence* of items.
- For health assessment *absence* of symptoms also important.
- In calibration, both presence and absence included ('doubling').
- Unsupervised algorithm as supervised (Fürnkranz et al., 2012).

Rule data base

- Standard analysis focuses on *presence* of items.
- For health assessment *absence* of symptoms also important.
- In calibration, both presence and absence included ('doubling').
- Unsupervised algorithm as supervised (Fürnkranz et al., 2012).
- Note that all variables are binary.

Item selection

- What item is most informative for criterion?
- Several statistics may be used:
  - Correlation (ϕ).
  - Odds-ratio.
  - Entropy.
  - .
- Each requires 2×2 table.

Required 2×2 table

Diagnosis

	$Y = 0$	$Y = 1$	
$X_j = 0, x_{i_1}, \ldots, x_{i_{k-1}}$	π_{00}	π_{01}	$1 - Q$
$X_j = 1, x_{i_1}, \ldots, x_{i_{k-1}}$	π_{10}	π_{11}	Q
	$1 - P$	P	

Cell probabilities obtainable from statistics

lhs		rhs	support	confidence	lift	count
{Back_R_Ankle}	=>	{crit0}	0.4478873	0.5530435	0.9647687	318
{Back_L_Knee}	=>	{crit0}	0.4309859	0.5303293	0.9251445	306
{Back_L_Wrist}	=>	{crit0}	0.4126761	0.5077990	0.8858409	293
{Back_R_Knee}	=>	{crit0}	0.4408451	0.5378007	0.9381781	313
{Back_L_Ankle}	=>	{crit0}	0.4492958	0.5452991	0.9512590	319
{Back_R_Wrist}	=>	{crit0}	0.4239437	0.5136519	0.8960512	301
{Back_L_Hip}	=>	{crit0}	0.4380282	0.5262267	0.9179877	311
{Back_R_Hip}	=>	{crit0}	0.4492958	0.5370370	0.9368459	319
{Front_L_Elbow}	=>	{crit0}	0.4605634	0.5351882	0.9336207	327
{Front_R_Elbow}	=>	{crit0}	0.4633803	0.5384615	0.9393309	329

Test score and stopping rule

Estimate of criterion probability after k items:

- $P(Y = 1 | x_{i_1}, \ldots, x_{i_k})$.
- $P(Y = 0) = 1 - P(Y = 1)$.

Test score and stopping rule

Estimate of criterion probability after k items:

- $P(Y = 1|x_{i_1}, \ldots, x_{i_k})$.
- $P(Y = 0) = 1 - P(Y = 1)$.

Stopping rule:

- Set required certainty γ (e.g. 0.95).
- Stop if $P(Y = 1) > \gamma$ or if $P(Y = 1) < 1 - \gamma$.

Pseudo-code for training phase

1: Data.0 $\leftarrow$ Combine item set and criterion $=$ 0 into data base
Req.0 $\leftarrow$ Set requirements for rule quality in Data.0
Results.0 $\leftarrow$ Run `apriori` on Data.0 using Req.0
Rules.0 $\leftarrow$ Rules from Results.0 with criterion $=$ 0 as consequent
2: Data.1 $\leftarrow$ Combine item set and criterion $=$ 1 into data base
Req.1 $\leftarrow$ Set requirements for rule quality in Data.1
Results.1 $\leftarrow$ Run `apriori` on Data.1 using Req.1
Rules.1 $\leftarrow$ Rules from Results.1 with criterion $=$ 1 as consequent
3: Rules $\leftarrow$ Join Rules.0 and Rules.1

Pseudo-code for application phase

Require: Rules
Require: γ

PPV $\leftarrow$ 0
NPV $\leftarrow$ 0
Items.left $\leftarrow$ item set
Items.used $\leftarrow$ empty
while PPV$< \gamma$ **and** NPV$> 1 - \gamma$ **and** cardinality of items.left > 0 **do**
    Pattern $\leftarrow$ response pattern to Items.used
    Rules.s $\leftarrow$ rules with Pattern as sub pattern and cardinality + 1
    if Rules.s is not empty **then**
        Select item with highest statistic.
    else if Rules.s is empty **then**
        Select item randomly
    end if
    Administer item
    Remove item from Items.left
    Add item to Items.used
    PPV $\leftarrow P(Y = 1)$ given response pattern
    NPV $\leftarrow P(Y = 0)$ given response pattern
end while

Output: PPV, NPV, Items.used, Pattern.

Synthetic data

Prediction of criterion score using 17 symptoms

```
$prob.pos
            [,1]
 [1,] 0.07939914
 [2,] 0.08928571
 [3,] 0.08406114
 [4,] 0.12000000
 [5,] 0.14285714
 [6,] 0.13333333
 [7,]        NaN
 [8,]        NaN
 [9,]        NaN
[10,]        NaN
[11,]        NaN
[12,]        NaN
[13,]        NaN
[14,]        NaN
[15,]        NaN
[16,]        NaN
[17,]        NaN
```

```
$`in.basket`
  "n.MSA_Q_08" "MSA_Q_01"   "MSA_Q_02"
  "MSA_Q_15"   "MSA_Q_16"   "MSA_Q_06"
  "MSA_Q_03"   "MSA_Q_04"   "n.MSA_Q_05"
  "MSA_Q_07"   "MSA_Q_09"   "MSA_Q_10"
  "MSA_Q_11"   "n.MSA_Q_12" "n.MSA_Q_13"
  "MSA_Q_14"   "n.MSA_Q_17"
```

What did I learn?

- apriori may have interesting features for adaptive testing.

What did I learn?

- apriori may have interesting features for adaptive testing.
- But: What to do with infrequent response patterns?

What did I learn?

- apriori may have interesting features for adaptive testing.
- But: What to do with infrequent response patterns?
- But: Didn't I just program a classification tree?

What did I learn?

- apriori may have interesting features for adaptive testing.
- But: What to do with infrequent response patterns?
- But: Didn't I just program a classification tree?
- Perhaps focus on unsupervised part:

What did I learn?

- apriori may have interesting features for adaptive testing.
- But: What to do with infrequent response patterns?
- But: Didn't I just program a classification tree?
- Perhaps focus on unsupervised part:
  - Look for many absents or presents of symptoms.

What did I learn?

- apriori may have interesting features for adaptive testing.
- But: What to do with infrequent response patterns?
- But: Didn't I just program a classification tree?
- Perhaps focus on unsupervised part:
  - Look for many absents or presents of symptoms.
  - Combine with Stochastic Curtailment.

What did I learn?

- apriori may have interesting features for adaptive testing.
- But: What to do with infrequent response patterns?
- But: Didn't I just program a classification tree?
- Perhaps focus on unsupervised part:
  - Look for many absents or presents of symptoms.
  - Combine with Stochastic Curtailment.
- I have to re-evaluate.

What did I learn?

- apriori may have interesting features for adaptive testing.
- But: What to do with infrequent response patterns?
- But: Didn't I just program a classification tree?
- Perhaps focus on unsupervised part:
  - Look for many absents or presents of symptoms.
  - Combine with Stochastic Curtailment.
- I have to re-evaluate.
- Do you have suggestions?

Thanks for your attention!

n.smits@uva.nl

Butcher, J. N., Keller, L. S., & Bacon, S. F. (1985). Current developments and future directions in computerized personality assessment. *Journal of Consulting and Clinical Psychology*, *53*(6), 803–815.

Finkelman, M. D., Kulich, R. J., Zoukhri, D., Smits, N., & Butler, S. F. (2013). Shortening the current opioid misuse measure via computer-based testing: a retrospective proof-of-concept study. *BMC Medical Research Methodology*, *13*(1), 126.

Finkelman, M. D., Smits, N., Kim, W., & Riley, B. (2012). Curtailment and stochastic curtailment to shorten the CES-D. *Applied Psychological Measurement*, *36*, 632–658.

Fokkema, M., Smits, N., Kelderman, H., Carlier, I. V. E., & van Hemert, A. M. (2014). Combining decision trees and stochastic curtailment for assessment length reduction of test batteries used for classification. *Applied*

Psychological Measurement, *38*, 3–17.

Fürnkranz, J., Gamberger, D., & Lavrač, N. (2012). *Foundations of rule learning*. London: Springer.

Oosterveld, P., Vorst, H. C. M., & Smits, N. (2019). Methods for questionnaire design: A taxonomy linking procedures to test goals. *Quality of Life Research*, *28*, 2501–2512. doi: 10.1007/s11136-019-02209-6

Smits, N., & Finkelman, M. D. (2015). Shortening the PHQ-9: A proof of principle study of utilizing stochastic curtailment as a method for constructing ultra-short screening instruments. *General Hospital Psychiatry*, *37*(5), 464–469.

Smits, N., van der Ark, L. A., & Conijn, J. M. (2018). Measurement versus prediction in the construction of patient-reported outcome questionnaires: Can we have our cake and eat it? *Quality of Life Research*, *27*, 1673–1682. doi: 10.1007/s11136-017-1720-4