An assistive handwashing system with emotional intelligence

Luyuan Lin

University of Waterloo

Supervisor:
Jesse Hoey

July 17, 2014

Overview

1. Problem Statement
   - Motivation
   - Objectives
2. Basic Concepts
   - Affect Control Theory (ACT)
   - Partially Observable Markov Decision Process (POMDP)
   - BayesACT
3. Solution
   - Selection based on battery depletion rate
   - Mobile battery consumption model
   - Missingness encoding
4. Experiment and Evaluation
   - Learn the treatment policy
   - Match with propensity score
   - Analyse the matching results
5. Conclusions and Future Plan

Problem Statement - Motivation

The COACH system:

- is an assistive system helping with an elder's daily activities
- monitors a user washing his/her hands and when needed
- detects when the user has lost track of what he/she is doing
- displays a prerecorded assistive prompt
- works well for some persons, but not as well for others

Problem Statement - Motivation

The COACH system:

- is an assistive system helping with an elder's daily activities
- monitors a user washing his/her hands and when needed
- detects when the user has lost track of what he/she is doing
- displays a prerecorded assistive prompt
- works well for some persons, but not as well for others

Using Emotional Intelligence in Assitive Systems

- recognize affective states
- generate affective signals
- study human emotions
- computationally model affective HCIs

Problem Statement - Objectives

To augment the COACH system with an emotional reasoning engine based on BayesACT so that the augmented system:

- is designed in a portable and extensible way
- runs in real-time from the perspective of the user group
- provides at least a level of functional assistance of as high quality as the COACH
- is able to tune the prompts in some way according to the emotional state of a user

Note: The last objective is ill-defined, as the question of how exactly tuning prompts to users will be most effective is not clear at this point.

Concepts - ACT

Affect Control Theory (ACT)

- represents emotions as vectors that represent evaluation (E), potency (P), and activity (A) respectively

Concepts - ACT

Affect Control Theory (ACT)

- represents emotions as vectors that represent evaluation (E), potency (P), and activity (A) respectively
- describes social events by an Actor-Behaviour-Object (ABO) grammar

Concepts - ACT

Affect Control Theory (ACT)

- represents emotions as vectors that represent evaluation (E), potency (P), and activity (A) respectively
- describes social events by an Actor-Behaviour-Object (ABO) grammar
- "fundamentals" of identities and behaviours; shared between people within a same culture

Concepts - ACT

Affect Control Theory (ACT)

- represents emotions as vectors that represent evaluation (E), potency (P), and activity (A) respectively
- describes social events by an Actor-Behaviour-Object (ABO) grammar
- "fundamentals" of identities and behaviours; shared between people within a same culture
- "transient impressions": emotional feelings of people evoked by a specific event

Concepts - ACT

Affect Control Theory (ACT)

- represents emotions as vectors that represent evaluation (E), potency (P), and activity (A) respectively
- describes social events by an Actor-Behaviour-Object (ABO) grammar
- "fundamentals" of identities and behaviours; shared between people within a same culture
- "transient impressions": emotional feelings of people evoked by a specific event

The ACT Principal

Actors work to experience transient impressions that are consistent with their fundamental sentiments.

Concepts - POMDP

Markov Decision Process (MDP)

- agent makes sequential decisions on discrete time
- $< S, A, T(s^{t+1}, s^t, a^t), R(s^{t+1}, s^t, a^t) >$
- maximize long term rewards
- policy - mapping from state to action

Concepts - POMDP

Markov Decision Process (MDP)

- agent makes sequential decisions on discrete time
- $< S, A, T(s^{t+1}, s^t, a^t), R(s^{t+1}, s^t, a^t) >$
- maximize long term rewards
- policy - mapping from state to action

Partially Observable MDP (POMDP)

- extended from MDP, involves observations
- maximize long term rewards based on history (actions and observations)
- computationally harder

Concepts - BayesACT

Markov Decision Process (MDP)

- agent makes sequential decisions on discrete time
- $< S, A, T(s^{t+1}, s^t, a^t), R(s^{t+1}, s^t, a^t) >$
- maximize long term rewards
- policy - mapping from state to action

Concepts - BayesACT

Markov Decision Process (MDP)

- agent makes sequential decisions on discrete time
- $< S, A, T(s^{t+1}, s^t, a^t), R(s^{t+1}, s^t, a^t) >$
- maximize long term rewards
- policy - mapping from state to action

Partially Observable MDP (POMDP)

- extended from MDP, involves observations
- maximize long term rewards based on history (actions and observations)
- computationally harder

Selection Criteria

For the good of mobile user experience, what is the most concerning?

Selection Criteria

For the good of mobile user experience, what is the most concerning?

Concerns

- slow
- weak signal strength
- crashing down
- short battery life (top gripe)

Selection Criteria

For the good of mobile user experience, what is the most concerning?

Concerns

- slow
- weak signal strength
- crashing down
- short battery life (top gripe)

Focus on battery life

- battery life is visible (opposed to CPU, Memory)
- simplified for research purpose
- battery depletion rate (ranking of implementations as suggestion)

How to obtain battery depletion rate

Pure experiments

- record battery depletion rate for every implementation candidates
- low-efficient

How to obtain battery depletion rate

Pure experiments

- record battery depletion rate for every implementation candidates
- low-efficient

Experiments & Estimation

- do some benchmark experiments
- estimate a battery depletion rate based on benchmark and other information

Mobile battery consumption model

Battery consumption comes from different components

- signal standby
- screen display
- CPU
- WIFI/3G
- sensors

Mobile battery consumption model

Battery consumption comes from different components

- signal standby
- screen display
- CPU
- WIFI/3G
- sensors

The usage of these components can be translated into battery consumption

Mobile battery consumption model

Battery consumption comes from different components

- signal standby
- screen display
- CPU
- WIFI/3G
- sensors

The usage of these components can be translated into battery consumption

We consider CPU and WIFI (ignore other three)

Mobile battery consumption model details

Suppose the POMDP makes decision every T (actual execution time is t, and the rest $T - t$ is idle)

- $r_T = r_{CPU} * t/T + r_{Base}(T)$ (mobile only, WIFI off)
- $r_T = r_{WIFICom} * t/T + r_{WIFIIdle} * (T - t)/T + r_{Base}(T)$ (mobile/server)

Mobile battery consumption model details

Suppose the POMDP makes decision every T (actual execution time is t, and the rest $T - t$ is idle)

- $r_T = r_{CPU} * t/T + r_{Base}(T)$ (mobile only, WIFI off)
- $r_T = r_{WIFICom} * t/T + r_{WIFIIdle} * (T - t)/T + r_{Base}(T)$ (mobile/server)

Mobile battery consumption model details

Benchmark: $< r_{CPU}, r_{WIFICom}, r_{WIFIIdle}, r_{Base}(T) >$

r_{Zero}: BDR of mobile device doing nothing

r_{CPU}: BDR of full cycle CPU usage minus r_{Zero}

$r_{WIFICom}$: BDR of continuous WIFI communication minus r_{Zero}

$r_{WIFIIdle}$: BDR of WIFI on minus r_{Zero}

$r_{Base}(T)$:

Mobile battery consumption model example

For a particular POMDP problem, given a set of implementation candidates implm1, implm2, implm3, ..., we examine the average BDR on three typical time interval T1, T2, T3:

- obtain benchmark $< r_{CPU}, r_{WIFICom}, r_{WIFIIdle}, r_{Base}(T) >$
- record average actual execution time t for each implementation candidate
- apply to the formula above, get r_{T_1}, r_{T_2}, r_{T_3} for each implementation candidate
- rank implementations based on their $\frac{r_{T_1}+r_{T_2}+r_{T_3}}{3}$

Mobile battery consumption model advantages

Using mobile battery consumption model to estimate BDR is fast

- a lot of implementation candidates to rank, only minutes for each (opposed to actual battery experiment for each)
- benchmark is device-determined, independent from POMDP problem
- benchmark might be shared among similar kind of devices (have no/little effects on ranking)

Data is big

Analysing a huge and complex dataset is both exciting and challenging:

Data is big

Analysing a huge and complex dataset is both exciting and challenging:

- It can make information transparent and usable at a much higher frequency.
- It contains lots of new knowledge for us to discover.

Data is big

Analysing a huge and complex dataset is both exciting and challenging:

- It can make information transparent and usable at a much higher frequency.
- It contains lots of new knowledge for us to discover.
- But, it would bring really high dimensionality to models, complex to find causal relationships hidden behind the data.

Data is big

Analysing a huge and complex dataset is both exciting and challenging:

- It can make information transparent and usable at a much higher frequency.
- It contains lots of new knowledge for us to discover.
- But, it would bring really high dimensionality to models, complex to find causal relationships hidden behind the data.
- Especially from dataset with a lot of missingness.

Data missingness

Features	NA rate	Ventilated NA	Non-Ventilated NA
max.pH	60%	24%	77%
min.pH	67%	34%	83%
max.pCO2	60%	24%	77%
min.pCO2	67%	34%	83%
WorstComaStatus	79%	85%	76%
...	...	...	...
Outcomes	**0 rate**	**Ventilated 0**	**Non-Ventilated 0**
DeltaPOPC*	86%	77%	90%
DeltaPCPC*	93%	85%	97%

Table: Missing (NA) rate for important features

*POPC: Pediatric Overall Performance Category *PCPC: Pediatric Cerebral Performance Category

Data missingness

Features	NA rate	Ventilated NA	Non-Ventilated NA
max.pH	60%	24%	77%
min.pH	67%	34%	83%
max.pCO2	60%	24%	77%
min.pCO2	67%	34%	83%
WorstComaStatus	79%	85%	76%
...	...	...	...
Outcomes	**0 rate**	**Ventilated 0**	**Non-Ventilated 0**
DeltaPOPC*	86%	77%	90%
DeltaPCPC*	93%	85%	97%

Table: Missing (NA) rate for important features

*POPC: Pediatric Overall Performance Category *PCPC: Pediatric Cerebral Performance Category

Differences between ventilated and non-ventilated missing rate might due to standard medical procedures, and is likely to remain the same in the future.

Shall we impute the data?

Data imputation: try to predict and fill in missing values from observed values

Missingness encoding

Because data are missing for a reason:

Features	**NA rate**	**Ventilated NA**	**Non-Ventilated NA**
PhHigh	60%	24%	77%
PhLow	67%	34%	83%
Pco2High	60%	24%	77%
Pco2Low	67%	34%	83%
WorstComaStatus	79%	85%	76%
...	...	...	...

Missingness encoding

Based on the assumption that missingness distribution will still be the same in the future, we take two encoding approaches for two different models:

- for regression models, we encode each NA as "0", add a dummy variable column for each feature indicating missingness (1=missing, 0=not missing)
- for tree models, we encode each NA as mean value of the column, add a dummy variable column for each feature indicating missingness (1=missing, 0=not missing)

Causal Inference

correlation.png

Causal Inference

We need to evaluate the effect of ventilator on PICU patient's outcome (mortality), so causal inference model is necessary.

Causal Inference

We need to evaluate the effect of ventilator on PICU patient's outcome (mortality), so causal inference model is necessary.

Ideally, with randomized trial data...

Causal Inference

We need to evaluate the effect of ventilator on PICU patient's outcome (mortality), so causal inference model is necessary.

Ideally, with randomized trial data...

- ventilators are assigned randomly
- treated group and control group have identical feature distribution, no confounding

Causal Inference

We need to evaluate the effect of ventilator on PICU patient's outcome (mortality), so causal inference model is necessary.

Ideally, with randomized trial data...

- ventilators are assigned randomly
- treated group and control group have identical feature distribution, no confounding
- we can compare outcome of treated patients with outcome of control patients to identify the effect of mechanical ventilation
- we can also use regression to find out which patients benefit the most from mechanical ventilation

Causal Inference

We need to evaluate the effect of ventilator on PICU patient's outcome (mortality), so causal inference model is necessary.

Ideally, with randomized trial data...

- ventilators are assigned randomly
- treated group and control group have identical feature distribution, no confounding
- we can compare outcome of treated patients with outcome of control patients to identify the effect of mechanical ventilation
- we can also use regression to find out which patients benefit the most from mechanical ventilation

However, in reality...

- ventilators were assigned according to clinical judgement, since randomization is not ethical
- treated group and control group have very different feature distributions

Causal Inference

We need to evaluate the effect of ventilator on PICU patient's outcome (mortality), so causal inference model is necessary.

Ideally, with randomized trial data...

- ventilators are assigned randomly
- treated group and control group have identical feature distribution, no confounding
- we can compare outcome of treated patients with outcome of control patients to identify the effect of mechanical ventilation
- we can also use regression to find out which patients benefit the most from mechanical ventilation

However, in reality...

- ventilators were assigned according to clinical judgement, since randomization is not ethical
- treated group and control group have very different feature distributions —ventilated patients are "sicker"

Causal Inference

We need to evaluate the effect of ventilator on PICU patient's outcome (mortality), so causal inference model is necessary.

Ideally, with randomized trial data...

- ventilators are assigned randomly
- treated group and control group have identical feature distribution, no confounding
- we can compare outcome of treated patients with outcome of control patients to identify the effect of mechanical ventilation
- we can also use regression to find out which patients benefit the most from mechanical ventilation

However, in reality...

- ventilators were assigned according to clinical judgement, since randomization is not ethical
- treated group and control group have very different feature distributions —ventilated patients are "sicker"
- if we simply compare their outcomes, ventilated patients mortality rate seems to be 7% higher than non-ventilated patients, but it does not mean giving a ventilator is bad

Causal Inference

We need to evaluate the effect of ventilator on PICU patient's outcome (mortality), so causal inference model is necessary.

Ideally, with randomized trial data...

- ventilators are assigned randomly
- treated group and control group have identical feature distribution, no confounding
- we can compare outcome of treated patients with outcome of control patients to identify the effect of mechanical ventilation
- we can also use regression to find out which patients benefit the most from mechanical ventilation

However, in reality...

- ventilators were assigned according to clinical judgement, since randomization is not ethical
- treated group and control group have very different feature distributions —ventilated patients are "sicker"
- if we simply compare their outcomes, ventilated patients mortality rate seems to be 7% higher than non-ventilated patients, but it does not mean giving a ventilator is bad
- How to compare outcomes?

Match the patients

We need to match patients:

- Matching tries to find each treated individual (ventilated patient) a treated-patient-counterpart (matched control patient) in the control group
- This matched control patient should have the most similar feature values to the treated individual

Match the patients

We need to match patients:

- Matching tries to find each treated individual (ventilated patient) a treated-patient-counterpart (matched control patient) in the control group
- This matched control patient should have the most similar feature values to the treated individual
- The outcome of the matched control patient is used for the **counterfactual outcome** of the treated patient (the outcome had the treated patient not received a ventilator) according to the Rubin Causal Model
- One of the most popular and promising methods of matching is the use of Propensity Scores.

Propensity Score Matching

Propensity score: the conditional probability of receiving the treatment given an individual's features

This propensity score model is also known as the "treatment policy"

Propensity Score Matching

Propensity score: the conditional probability of receiving the treatment given an individual's features

This propensity score model is also known as the "treatment policy"

Use propensity score model to conduct matching:

1. use statistical/machine learning models to learn the treatment policy from the dataset (except outcome)
2. record the propensity score this treatment policy assigns for each patient
3. find each treated (ventilated) patient a match patient in the control (non-ventilated) group with the closest propensity score
4. a control individual might be matched more than once

Re-construct the treatment policy

We use logistic regression and decision tree model to learn the assignment policy from the data, but only logistic regression results are used in matching:

Logistic Regression Ventilator Treatment Policy Model

use ventilator assignment as model output, remove all the outcomes from feature set, use the rest of features (NA encoded as "0") as well as their missing indicators

```
Ventilator~vars+missingness
```

Decision Tree Ventilator Treatment Policy Model

classify ventilator assignment, remove outcome features, use the rest of mean-imputed features as well as their missing indicators:

```
Ventilator~vars+missingness
```

Decision Tree Ventilator Treat Policy Model

rpart08051.png

Logistic Regression Ventilator Treatment Policy Model

Feature	Estimate	Std.Error	Signif.Code
max.pH	5.7142666	0.312283	***
min.pH	-2.43043	0.297853	***
(Intercept)	2.1823757	0.1620187	***
origin20	1.9733549	0.1797177	***
dx15PimLowRiskYes	-1.8883596	0.3229063	***
dx16Prism1	1.823	0.3065	***
dx15Prism1	-1.4193727	0.3100689	***
NA.GlasgowComaScore1	1.7761392	0.0358459	***
NA.max.pH1	-0.9285517	0.1927179	***
PopcAdmit4	1.2742374	0.081754	***
NA.PupilReaction1	-0.827305	0.0466854	***
PcpcAdmit4	-0.7671332	0.0876789	***
...	...	...	...

Table: Logistic Regression Ventilator Treatment Policy Model
Sig.Code: *** p-value<0.001; ** 0.001<=p-value<0.01; * 0.01<=p-value<0.05

Matching results comparison

Model	Estimated Effect*
Without matching	0.06907023
Logistic regression matching	0.0029017

Table: Matching results comparison

*Average Estimated Effect of Mechanical Ventilator on Mortality among Ventilated Patients

Several thoughts

Estimated effect on mortality is not significantly different from zero

Several thoughts

Estimated effect on mortality is not significantly different from zero

Possibilities:

1. The counterfactual outcomes from matching are of poor quality

Several thoughts

Estimated effect on mortality is not significantly different from zero

Possibilities:

1. The counterfactual outcomes from matching are of poor quality
2. Ventilators have different effects among different sub-populations

Matching is not good enough?

Possibility 1: Matching (propensity score model) is not good enough?

Quality of Matching

Feature	var.ratio* before matching	var.ratio after matching	improved?
PhHigh	2.9134	1.1205	Yes
Pco2High	5.0744	0.69964	Yes
PupilReaction3	3.2278	1.4426	Yes
GlasgowComaScore	3.6439	0.8961	Yes
origin20	9.2555	0.70828	Yes
PopcAdmit4	1.8238	0.82998	Yes
originAnotherICU1	1.843	1.021	Yes
postop	1.0259	1.0691	No
...	...	...	...

Table: MatchBalance performance of Logistic Regression Propensity Score Model

*var.ratio before matching=var.treated/var.control

*var.ratio after matching=var.treated/var.matched

Quality of Matching

Feature	var.ratio* before matching	var.ratio after matching	improved?
PhHigh	2.9134	1.1205	Yes
Pco2High	5.0744	0.69964	Yes
PupilReaction3	3.2278	1.4426	Yes
GlasgowComaScore	3.6439	0.8961	Yes
origin20	9.2555	0.70828	Yes
PopcAdmit4	1.8238	0.82998	Yes
originAnotherICU1	1.843	1.021	Yes
postop	1.0259	1.0691	No
...	...	...	...

Table: MatchBalance performance of Logistic Regression Propensity Score Model

*var.ratio before matching=var.treated/var.control
*var.ratio after matching=var.treated/var.matched

Features are not perfectly balanced, but matching improves most of them.

Histogram of regression treatment policy model

testhist.pdf

Ventilator does not always save?

Possibility 2: Ventilators have different effects on different patients?

Who might benefit the most from a ventilator?

1. According to the Rubin Causal Model, we construct a new binary outcome named "helped":

   ```
   if treated.Mortality < control.Mortality: helped = 1;
   else: helped = 0
   ```

Who might benefit the most from a ventilator?

1. According to the Rubin Causal Model, we construct a new binary outcome named "helped":

   `if treated.Mortality < control.Mortality: helped = 1;`
   `else: helped = 0`

2. Use decision tree model to learn and classify this new outcome from all features among treated patients

   `helped~vars+missingness (among treated patients)`

Decision Tree "helped" Model:

finaltree.png

Conclusions

- Casual reasoning is necessary to obtain meaningful results in this setting
- Though the missing of important features and their measurement time make causal inference very challenging, we are still cautiously optimistic that matching corrects for imbalances between the treated and control groups
- Currently we believe that patients who have their pCO2 measured and a low Glasgow Coma Score are most likely to be helped by mechanical ventilation. We are interested in discussing with clinicians why this might be the case.

Future Plans

- In the future, we will conduct a simulation study using synthetic data similar to our real data to assess the performance of matching when we know the ground truth.
- We will investigate whether some counterfactual outcomes of treated patients are in fact known, and do not need to be matched: For example, some ventilated patients would certainly die without a ventilator.
- We aim to develop a policy for ventilator triage.

Future Plans

- In the future, we will conduct a simulation study using synthetic data similar to our real data to assess the performance of matching when we know the ground truth.
- We will investigate whether some counterfactual outcomes of treated patients are in fact known, and do not need to be matched: For example, some ventilated patients would certainly die without a ventilator.
- We aim to develop a policy for ventilator triage.
- We will try the model on the new dataset!

Acknowledgements

- We would like to thank Virtual PICU Systems (VPS) and Children's Hospital of Los Angeles (CHLA) for collecting and generously providing the data for our project.
- We also would like to thank Professor Dr. Randal Wetzel and Mr. David Kale for their kindly help and advices to our project.
- We also acknowledge support from the Natural Sciences and Engineering Research Council (NSERC) of Canada.

References

[1] Brookhart, M Alan, Schneeweiss, Sebastian, Rothman, Kenneth J, Glynn, Robert J, Avorn, Jerry, and Sturmer, Til. Variable selection for propensity score models. American journal of epidemiology, 163(12):1149-1156, 2006.

[2] Garg, Amit X, Adhikari, Neill KJ, McDonald, Heather, Rosas-Arellano, M Patricia, Devereaux, PJ, Beyene, Joseph, Sam, Justina, and Haynes, R Brian. Effects of computerized clinical decision support systems on practitioner performance and patient outcomes. JAMA: the journal of the American Medical Association, 293(10):1223-1238, 2005.

[3] Greenland, Sander. Quantifying biases in causal models: classical confounding vs. collider stratication bias. Epidemiology, 14(3):300-306, 2003.

[4] Ihaka, Ross. R: Past and future history. COMPUTING SCIENCE AND STATISTICS, pp. 392-396, 1998.

[5] Kale, David. Quick and dirty data dictionary for vps data. Technical report, Children's Hospital Los Angeles, Los Angeles,CA, 2012.

[6] Pollack, Murray M, Patel, Kantilal M, and Ruttimann, Urs E. Prism iii: an updated pediatric risk of mortality score. Critical care medicine, 24(5):743-752, 1996.

[7] Rosenbaum, Paul R and Rubin, Donald B. The central role of the propensity score in observational studies for causal e ects. Biometrika, 70(1):41-55, 1983.

[8] Rubin, Donald B. Estimating causal e ects of treatments in randomized and nonrandomized studies. Journal of educational Psychology, 66(5): 688-701, 1974.

[9] Rubin, Donald B and Thomas, Neal. Matching using estimated propensity scores: relating theory to practice. Biometrics, pp. 249-264, 1996.

[10] Stuart, Elizabeth A. Matching methods for causal inference: A review and a look forward. Statistical science: a review journal of the Institute of Mathematical Statistics, 25(1):1, 2010.

The end

Thank you!

- Questions?
- Comments?