Leveraging Routine Behavior and Contextually-Filtered Features for Depression Detection among College Students

XUHAI XU, University of Washington
PRERNA CHIKERSAL, AFSANEH DORYAB, DANIELLA K. VILLALBA, and JANINE M. DUTCHER, Carnegie Mellon University
MICHAEL J. TUMMINIA, University of Pittsburgh
TIM ALTHOFF, University of Washington
SHELDON COHEN, KASEY G. CRESWELL, and J. DAVID CRESWELL, Carnegie Mellon University
JENNIFER MANKOFF and ANIND K. DEY, University of Washington

The rate of depression in college students is rising, which is known to increase suicide risk, lower academic performance and double the likelihood of dropping out of school. Existing work on finding relationships between passively sensed behavior and depression, as well as detecting depression, mainly derives relevant *unimodal features* from a single sensor. However, co-occurrence of values in multiple sensors may provide better features, because such features can describe behavior *in context*. We present a new method to extract *contextually filtered* features from passively collected, time-series mobile data *via* association rule mining. After calculating traditional unimodal features from the data, we extract rules that relate unimodal features to each other using association rule mining. We extract rules from each class separately (*e.g.*, depression vs. non-depression). We introduce a new metric to select a subset of rules that distinguish between the two classes. From these rules, which capture the relationship between multiple unimodal features, we automatically extract *contextually filtered features*. These features are then fed into a traditional machine learning pipeline to detect the class of interest (in our case, depression), defined by whether a student has a high BDI-II score at the end of the semester. The behavior rules generated by our methods are highly interpretable representations of differences between classes. Our best model uses contextually-filtered features to significantly outperform a standard model that uses only unimodal features, by an average of 9.7% across a variety of metrics. We further verified the generalizability of our approach on a second dataset, and achieved very similar results.

CCS Concepts: • **Human-centered computing Ubiquitous and mobile computing**; • **Applied computing Life and medical sciences**.

Additional Key Words and Phrases: Behavior mining, Passive sensing, Depression detection, Association rule mining

ACM Reference Format:
Xuhai Xu, Prerna Chikersal, Afsaneh Doryab, Daniella K. Villalba, Janine M. Dutcher, Michael J. Tumminia, Tim Althoff, Sheldon Cohen, Kasey G. Creswell, J. David Creswell, Jennifer Mankoff, and Anind K. Dey. 2019. Leveraging Routine Behavior and Contextually-Filtered Features for Depression Detection among College Students. *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.* 3, 3, Article 116 (September 2019), 33 pages. https://doi.org/10.1145/3351274

116

Authors' addresses: Xuhai Xu, xuhaixu@uw.edu, University of Washington, 1410 NE Campus Parkway, Seattle, WA, 98195; Prerna Chikersal; Afsaneh Doryab; Daniella K. Villalba; Janine M. Dutcher, Carnegie Mellon University, Pittsburgh, PA, 15289; Michael J. Tumminia, University of Pittsburgh, Pittsburgh, PA, 15260; Tim Althoff, University of Washington; Sheldon Cohen; Kasey G. Creswell; J. David Creswell, Carnegie Mellon University; Jennifer Mankoff; Anind K. Dey, University of Washington.

2474-9567/2019/9-ART116 $15.00
https://doi.org/10.1145/3351274

1. Leveraging Routine Behavior and Contextually-Filtered Features for Depression Detection among College Students (12 slides)

Problem:

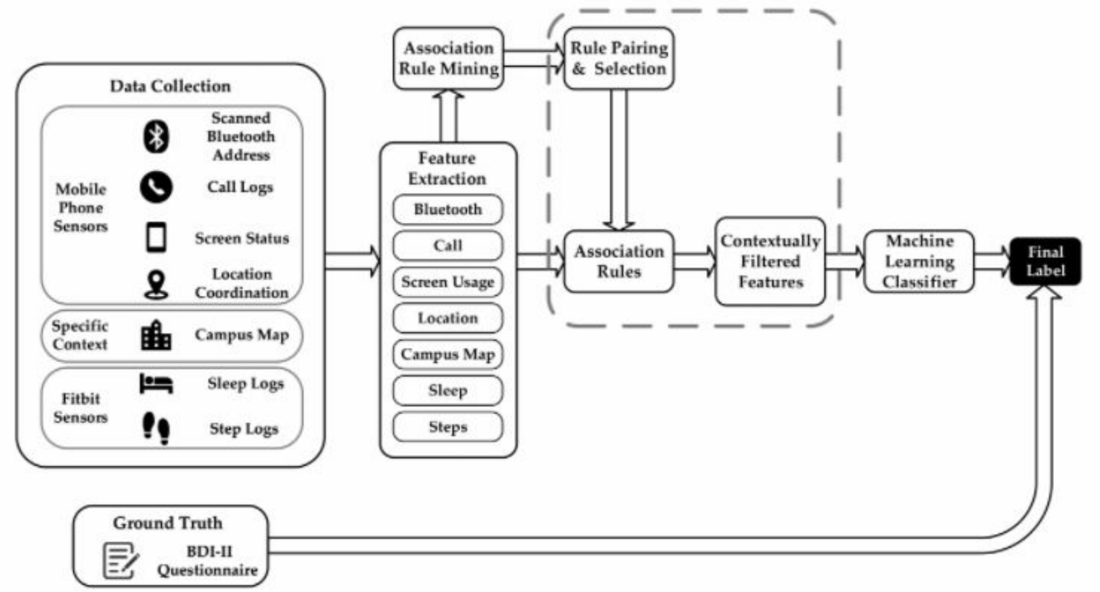
- Depression detection (DD) can mitigate or prevent its negative consequences (e.g., cognitive impairment)

Background:

- Previous work uses *unimodal features* where each sensor channel is used as a separate feature

Motivation:

- Co-occurrence of values in multiple sensors may provide better features for DD by describing behavior *in context*



1. Feature Engineering: unimodal features vs. contextually filtered features

Feature Extraction

- *Unimodal features*
  - how-often student ‘sitting in the classroom’ ***during a day***
  - how-often student ‘interacting w/ their phone’ ***during a day***
- *Contextual filtered features*
  - how-often student ‘interacting w/ their phone ***when*** they are sitting in the classroom’

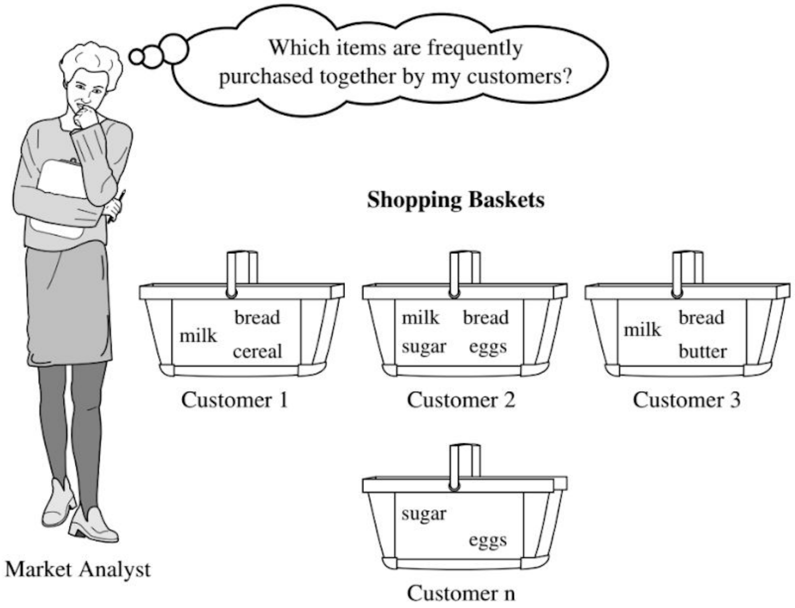
Contextual filtering with ARM?

- ARM decides which features to use for contextual filtering (X) and which feature to use for aggregation (Y) where {X} => {Y}
  - *Context*: days in which student is sitting in a classroom
  - *Feature for aggregation*: how often student interacted w/ their phone in a day for a given *context*

[1] Wahle, Fabian, et al. "Mobile sensing and support for people with depression: a pilot trial in the wild." *JMIR mHealth and uHealth* (2016).
[2] Saeb, Sohrab, et al. "Mobile phone sensor correlates of depressive symptom severity in daily-life behavior: an exploratory study." *JMIR*(2015)

Frequent pattern (association rule) mining (1/2)

Association Rule Mining (ARM) is a technique to uncover how items are associated to each other [1]



Milk and bread are frequently purchased together

[1] Han, Jiawei, Jian Pei, and Micheline Kamber. *Data mining: concepts and techniques*. Elsevier, 2011.

Frequent pattern (association rule) mining (2/2)

Rule interestingness measures (with example):

- **Support{apple,beer}** = ⅜ 37.5% of the time apple & beer are purchased together
- **Confidence {apple}-> {beer}** = **¾.** So we are 75% certain that if user purchases apple, he will also purchase beer.

Support {🍎} = $\frac{4}{8}$

Confidence {🍎 → 🍺} = $\frac{\text{Support }\{🍎, 🍺\}}{\text{Support }\{🍎\}}$

Associated Rule Mining Steps:

1. *Find all frequent itemsets*:
   a. each of these itemsets will occur at least as frequently as a predetermined minimum support. E: {apple, bear}
2. *Generate strong association rules from the frequent itemsets*:
   a. these rules must satisfy minimum support and minimum confidence. E: {apple}=> {bear}

Transaction 1	🍎 🍺 🍚 🍗
Transaction 2	🍎 🍺 🍚
Transaction 3	🍎 🍺
Transaction 4	🍎 🍐
Transaction 5	🍼 🍺 🍚 🍗
Transaction 6	🍼 🍺 🍚
Transaction 7	🍼 🍺
Transaction 8	🍼 🍐

Table 1. Example Transactions (baskets)

[1] Han, Jiawei, Jian Pei, and Micheline Kamber. *Data mining: concepts and techniques*. Elsevier, 2011.

Feature Engineering: unimodal features vs. contextually filtered features

Feature Extraction

- *Unimodal features*
  - how-often student 'sitting in the classroom' ***during a day***
  - how-often student 'interacting w/ their phone' ***during a day***
- *Contextual filtered features*
  - how-often student 'interacting w/ their phone ***when*** they are sitting in the classroom'

Contextual filtering with ARM?

- ARM decides which features to use for contextual filtering (X) and which feature to use for aggregation (Y) where {X} => {Y}
  - *Context*: days in which student is sitting in a classroom
  - *Feature for aggregation*: how often student interacted w/ their phone in a day for a given *context*

[1] Wahle, Fabian, et al. "Mobile sensing and support for people with depression: a pilot trial in the wild." *JMIR mHealth and uHealth* (2016).
[2] Saeb, Sohrab, et al. "Mobile phone sensor correlates of depressive symptom severity in daily-life behavior: an exploratory study." *JMIR*(2015)

Data Collection: Study Design

Ground Truth Data Collection:

- At the beginning and end of semester, BDI-II self-report test for the presence and severity of depressive symptoms

Study	Days	Overall Number	Dropped out Number	Removed Number	Dataset Size	Pre-semester BDI-II Non-dep Grp	Pre-semester BDI-II Dep Grp	Post-semester BDI-II Non-dep Grp	Post-semester BDI-II Dep Grp
Phase I	106	188	28	22	138	114	24	81	57 (41.3%)
Phase II	113	267	31	24	212	-	-	136	76 (35.8%)

Data Collection

Sensor	**Source**	**Sampling**	**Information Being Aggregated into Features**	**Number of Samples Per-person**
Screen	AWARE	event-based	Number of unlocks per minute, total time with interaction, total time unlocked	39843.2 ± 22126.9
Call	AWARE	event-based	Number and duration of in-coming /out-going/missed calls	379.6 ± 275.8
Bluetooth	AWARE	1 per 10 minutes	Number of unique devices, number of scans of most/least frequent device	24579.0 ± 106960.9
Location	AWARE	1 per 10 minutes	GPS latitude, longitude, altitude	9692.8 ± 4444.2
Sleep	Fitbit	1 per minute	Asleep/restless/awake/unknown duration and onset	34963.6 ± 15630.6
Step	Fitbit	1 per 5 minutes	Number of steps	23390.6 ± 10197.7

Sensor data and information aggregated into fatures

Unimodal Feature Extraction [1]

1. **raw sensors data => *epochs*** -/ to capture behaviour pattern during different times of the day
    a. 8 epochs: Weekday/Weekend **x** (Night, morning, afternoon, evening): e.g., Weekday Night

2. **epochs => *daily-epoch* features:**
    a. 8 epochs in each day
    b. E: the number of phone calls on the morning of Tuesday, February 18, 2017

3. **daily-epoch features => *full features* within an epoch (424 features)**
    a. 212 daily epoch features // 212x2 (mean and standard deviation of daily epoch features) *full* features
    b. *E*: *the average number of calls* (weekday-morning epoch)
        i. calculated as an average over the daily-epoch feature which captures the number of calls made each weekday morning

[1] Wang, Rui, et al. "StudentLife: assessing mental health, academic performance and behavioral trends of college students using smartphones." *UbiComp*. 2014.

Explanation: daily-epoch features: num_weeks*(5*4+2*4)

Weekday	epochs			
Week1. Monday	Night	Morning	Afternoon	Evening
Week1. Tue	Night	Morning	Afternoon	Evening
...				
Week9. Fri	Night	Morning	Afternoon	Evening

Weekend	epochs			
Week1. Sat	Night	Morning	Afternoon	Evening
Week1. Sun	Night	Morning	Afternoon	Evening
...				
Week9. Sun	Night	Morning	Afternoon	Evening

Explanation: daily-epoch features: num_weeks*(5*4+2*4)

Weekday	epochs			
Week1. Monday	Night	Morning	Afternoon	Evening
Week1. Tue	Night	Morning	Afternoon	Evening
...				
Week9. Fri	Night	Morning	Afternoon	Evening

Weekend	epochs			
Week1. Sat	Night	Morning	Afternoon	Evening
Week1. Sun	Night	Morning	Afternoon	Evening
...				
Week9. Sun	Night	Morning	Afternoon	Evening

Each daily-epoch has 212 features consisting of

- aggregated features using
  - For sampled: Mean, max, min, std
  - For event-based: count, duration
- Special features
  - **Location**: Location variance, Total distance traveled, average/variance of speed, Circadiant movement
  - **Location clusters**: #significant places, #transitions between places, radius of gyration[19], %time spent at top-3 frequented clusters/moving/rarely visited locations, length of stay at clusters, location entropy
  - **Steps: active bouts, sedentary bouts**
  - **bluetooth**
- **Example daily-epoch feature**: the number of phone calls on the morning of Tuesday, Week 1(February 18, 2017)

Explanation: Aggregate daily-epoch features within epoch => (Unimodal) full features

Weekday	epochs			
Week1. Monday	Night	Morning	Afternoon	Evening
Week1. Tue	Night	Morning	Afternoon	Evening
...				
Week9. Fri	Night	Morning	Afternoon	Evening

Weekend	epochs			
Week1. Sat	Night	Morning	Afternoon	Evening
Week1. Sun	Night	Morning	Afternoon	Evening
...				
Week9. Sun	Night	Morning	Afternoon	Evening

$\left.\right\}$ Mean, Std

Daily-epoch features

Feat. ID	Feature	Weekday: Night	Weekday: Morning	Weekday: Afternoon	Weekday: Evening	Weekend: Night	Weekend: Morning	Weekend: Afternoon	Weekend: Evening
1	MEAN of #phone calls	0	0	2	3	2	1	2	0
2	STD of #phone calls								
...									
424	STD of ...								

Full features

Rule Mining, Selection, Multimodal Feature Extraction

- **Generate rules for each group(depression and non-depression) from unimodal features** - /normal ARM procedure
  - E: [X : {Staying at home, Low activity level} → Y : {Being asleep}]

- **Select T_{top} rules ranking metrics** (pls refer to paper for details)
  - Rules are ranked higher if they help to identify *important differences* between depression and non-depression group
    - Rules that are same between two groups but different *sup* and *conf* - /similar context in each group, but different behaviour (*sup & conf*)
    - Rules that are unique to only one group - / different context in each group (*sup*)

- **Obtain CF features for each rule in T_{top}**
  - Select all time periods or epochs, E_p for person p that fulfil the context ***X***
    - E: {Staying at home, low activity level} context is observed in 3 epochs:
      - Week 1. Mon/Morning
      - Week 3. Sun/Afternoon
      - Week 9. Sun/ Evening
  - Then extract *mean* and *std* of the features in ***Y*** for all epochs that ***X*** extended
    - *mean* and *std* of {Long phone call duration}

Pipeline for detecting depression

1. **Data Set Preparation**
   a. RuleGenerateSet of 50 people (20/30 in the depression/non-depression group)
   b. TrainTestSet of 88 people to train and test ML models (37/51 depression/non-depression group)
2. **Feature Selection[1]** -/ to reduce computational complexity
   a. There are 424 (212) *full features* in each of 8 epochs
   b. Using mutual information [1], top 181 features are selected among 8 x 424 features - / 181 Unimodal Features
      i. 23 top (mean & std.) features on average (Min = 18, Max = 29, 181 in total) among the 8 epochs
3. **Mapping selected (unimodal) full features back to daily-epoch features**
   a. ~20 (Min=15, Max=29) daily-epoch features in each epoch
4. **Feature Preparation before Rule Mining**
   a. features are recoded (discretized) into the three categories: low, moderate, high only for the purpose of ARM
5. **Rule Mining and Selecting**
   a. For each epoch, on average 16,000 (Min=4500, Max=26,000) rules are generated from discretized top daily-epoch features
   b. For each epoch, on average 13 (Min=6, Max=19) rules are selected using proposed RuleSelection -/ total of 105 rules
6. **Feature Extraction and Model training**
   a. an average of 17(Min=8, Max=23) contextually filtered features are extracted per-epoch -/ in total 137x2 (mean and std) features

[1] Quinlan, J. Ross. "Induction of decision trees." *Machine learning* (1986).

Contextually Filtered Features Lead to Higher Performing Machine Learning Models

Classification	Features	Accuracy	Precision	Recall	F1 Score
	Majority	0.579	0.579	1.000	0.734
	Best Single Feature	0.704	0.725	0.755	0.740
CPAR [95]	Class Association Rules	0.608	0.629	0.850	0.723
AdaBoost [35]	Unimodal Features	0.716	0.725	0.771	0.747
AdaBoost [35]	Contextually Filtered Features	0.807	0.765	**0.886**	0.821
AdaBoost [35]	Hybrid Features	**0.818**	**0.843**	0.843	**0.843**
	Performance Increase of Hybrid over Unimodal	*10.2%*	*11.8%*	*7.2%*	*9.6%*
			Average Increase:		***9.7%***

Insights

- **CFF provides better feature representation**
- **Computationally expensive**