# Assignment – Munmun De Choudhury

Assistant Professor, School of Interactive Computing

Georgia Tech

*NSF International Workshop on Dynamic Modeling of Health Behavior Change and Maintenance,*
*Sept 8-9, 2015, London, UK*

# Background

- Skyler is middle-aged, overweight and needs to manage intake of different types of food (e.g., snacking between meals) as well as control the overall calorie intake.
- Goal – two kinds of interventions (1) interventions for long-term behavior change; (2) just-in-time adaptive interventions.

# Intervening for behavior change

- These interventions need to:
  - identify stressors that lead to unhealthy eating habits; identify (mediating) attributes that lead to Skyler's improved self-efficacy over time and what can lead to their adherence to better dietary choices

# Just in time adaptive interventions

- These interventions need to:
  - identify the variety of cues Skyler may be exposed to immediately preceding a binge/snack eating episode
  - identify Skyler's predisposition to binge eating based on eating patterns, lifestyle and psychosocial context in a longer time frame
  - model and predict the likelihood of both of these cues precipitating an actual eating episode

# Main idea of the computational models for the two interventions – a hierarchical model based on the social cognitive theory

# Integration/Fusion of Data Sources

- We will obtain Skyler's consent to obtain the following diverse data capturing affective, behavioral, cognitive attributes that may act as precipitation cues
  - Baseline data about Skyler, e.g., body fat, BMI, sleeping patterns, working hours, mental health status, assessment of blood glucose levels whenever available.
  - Self-reported data (via a a starting survey, six EMAs everyday about dietary patterns, items eaten, nutrient intake, everyday experiences of events, stress and anxiety)
  - Wearable sensing data (physical activity, heart rate variability, respiration)
  - Social media – access to social capital, social support and nature of affect, ingestion behavior
    - Social media can capture the broader content and context of interactions with others

# Social media

- Facebook, Twitter and Instagram are of particular interest
- Health and well-being related support forums e.g., on Reddit
- Use the respective APIs (Application Programming Interfaces) of the above platforms for data collection

# Social media measures (affect)

- Levels and types of affect manifested in social media posts will be measured, including how frequently and intensely various affective states such as, positive affect (PA), negative affect (NA), anger, fatigue, hostility, sadness, joviality, anxiety, fear are expressed over time.
- We will leverage our prior work in inferring a distribution over each affective state via supervised learning, specifically a multi-class maximum entropy classifier

# Social media measures (affect)

- Emotional activation and dominance will also be assessed over all of the social media posts of Skyler.
- We will first use the ANEW lexicon to label a seed set of social media posts with their activation and dominance measures. Then applying Bayesian belief propagation over a co-occurrence graph of unigrams and bigrams in these labeled posts we will learn relationships between them.
- These learned relationships will be used to infer activation and dominance in unlabeled posts

# Social media measures (topics)

- Topics will be inferred by employing the Latent Dirichlet Allocation (LDA) topic model in the social media content.
- LDA is a generative model and represents content (posts) as mixtures of topics that spit out words with certain probabilities; hence it is suitable to measure subjective thematic elements of what student discuss on social media

# Social media measures (engagement)

- Engagement will be measured based on the frequency of shares of different content types, e.g., textual posts, images, videos, links.
- We will also measure diurnal posting patterns and the entropy of posting activity.

# Social media measures (network structure)

- Social network structural attributes will be inferred based on the content exchange patterns in the social media:
  - network density, clustering coefficient, embeddedness, network constraint, average node (in)-degree and (out)-degree, reciprocity, and number of connected components in social network of Skyler
- These structural attributes have been known to be indicative of greater access to social capital, and hence better health

# A Model of Behavioral Capability

- Inputs: Skyler's personality attributes; Skyler's recognition of the risks of poor eating habits; historical data (from social media, self-reports) on patterns of when, where and what they have been eating
  - Long-term data spanning months or about an year
- Output: Behavioral capability measure
- Model: Supervised learning framework like Support Vector Regression, or a regularized logistic regression model

# A Model of Self-efficacy

- Inputs: historical data on Skyler's susceptibility to binge/ snack eating; availability of social and emotional support; historical information on Skyler's mental wellness; self-regulation score; nature of expression (on social media, through EMAs) regarding stressful life events and experiences; general self-efficacy scale score
  - Long-term data spanning months or about an year
- Output: self-efficacy measure
- Model: A latent variable model like a Conditional Random Field (discretized self-efficacy "states") where "self-efficacy" is the latent state.

# Modeling Reinforcements

- Inputs: social media affect (valence and arousal), topics related to affect, social interactions, wearable sensing information (stress, presence of cues that may trigger eating), EMA information on food intake
  - Past few hours of data, including fine-grained, real-time temporal assessments
- Output: binary indicator of positive or negative reinforcement
- Model:
  - Gaussian Process Regression to handle continuous and discretized data together in a model – well-validated interpolation model that is helpful for performing inference using potentially uncertain data
  - A linear quadratic estimation model (a Kalman filter) for fusing the input data into determining Skyler's reinforcement type at a particular point in time

# Predicting likelihood of a binge episode

- Inputs:
  - Inputs and outcomes of the earlier three models
- Output: likelihood of binge eating
- Model:
  - Due to the mutual interdependencies and hierarchical relationships between the various cues we measure and the three different models of social cognitive theory, a nested multi-layer deep neural network is suitable for estimating Skyler's likelihood of a binge episode
    - Behavioral capability and self-efficacy are important aspects to determine whether triggering an intervention will be effective

# Open questions

- Is integrating multiple sources of data really useful? What are the breakpoints?
- How much training do the models need? What is a baseline model to compare our proposed model with?
- How do we evaluate if our models are helpful to Skyler in the immediate or long-term?
- How can feedback be obtained from Skyler to improve the models over time?
- How do we handle false positives?

# Open Questions

- What would the interventions look like? How would they integrate the models?
- How would the interventions be made? Smartphone, social network?
- How can these models be generalized to other individuals with diverse characteristics and distinct from Skyler?