Policy-based Recommendations in a Flow Choice Architecture

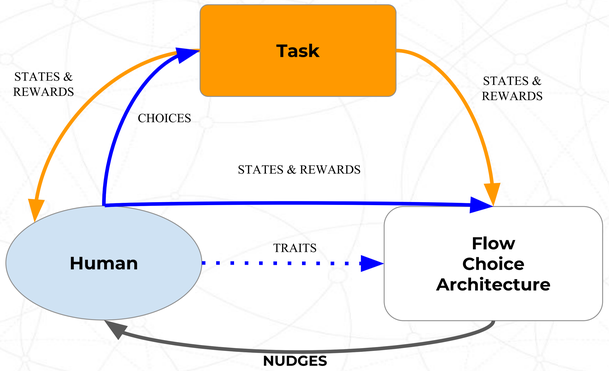
Troy R. Weekes (tweekes1998@my.fit.edu) and Thomas C. Eskridge (teskridge@fit.edu)

FLORIDA TECH

Purpose

Humans need to perform at high levels in knowledge work domains. This type of non-routine cognitive work requires considerable amounts of concentration and creativity to perform. Our research discusses policy-based algorithms that recommend appropriate nudges to induce and sustain flow performance within the dynamic and non-stationary environment of human knowledge work.

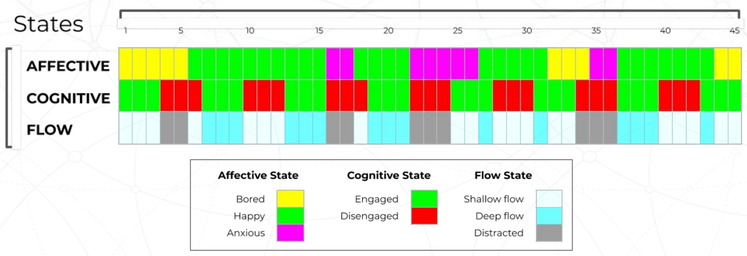
Background



The Flow Choice Architecture (FCA) observes human and task states and reward signals. It integrates that information with context in order to reduce uncertainty when recommending timely and task-relevant nudges.

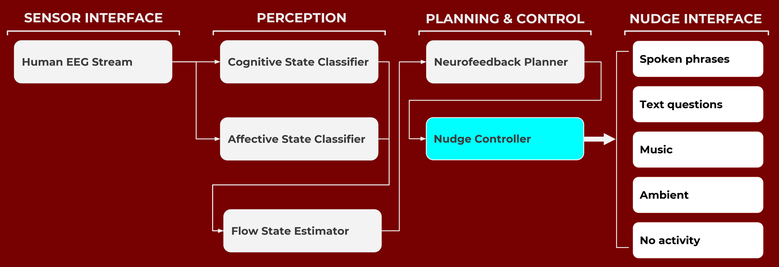
Methodology

A synthetic dataset of 2000 trials was used in this study. Trials were randomly sampled from 18 types of knowledge work scenarios. Each scenario type simulated an overarching affective state. In order to simulate different frequencies of distractions, perturbations were made to the cognitive state.



Human state classification during a 3-minute knowledge work task

Flow Choice Architecture



Recommending Effective Nudges

Human state transitions are confirmed when the flow estimate stabilizes in a new state. Human choices add personalized information to case representations. Contextual data such as current task, previous task and time of day are linked with human states. Self-reports are used to fine-tune the recommendation of nudges. Given the uncertain effects of nudges, FCA retrieves a set of alternative nudges for use in its policy-based exploration and exploitation algorithms.

Q-learning Algorithm

Let $a_1,\ldots, a_m$ be a set of nudges
Let $s_1,\ldots, s_n$ be a set of human states
Using Epsilon Greedy

Input: human states and set of nudges
Output: Nudges that maximize cumulative reward

Initialize Q(s, a) arbitrarily
Repeat (for each episode):
  Initialize s
  Repeat (for each step of episode):
    Choose a from s using Q policy based on epsilon
    Recommend nudge a, observe reward r, and state s′
    Update
      $Q(s, a) \leftarrow Q(s, a) + \alpha[r + \gamma \max_{a'} Q(s', a') - Q(s, a)]$
    $s \leftarrow s'$
  Until s is terminal i.e. the work session is complete.

Contextual Bandit Algorithm

Let $a_1,\ldots, a_m$ be a set of nudges
Given a context x_t
Let H be the history of {{context, nudge, reward),..}
Let θ be the model parameter
Using Thompson Sampling

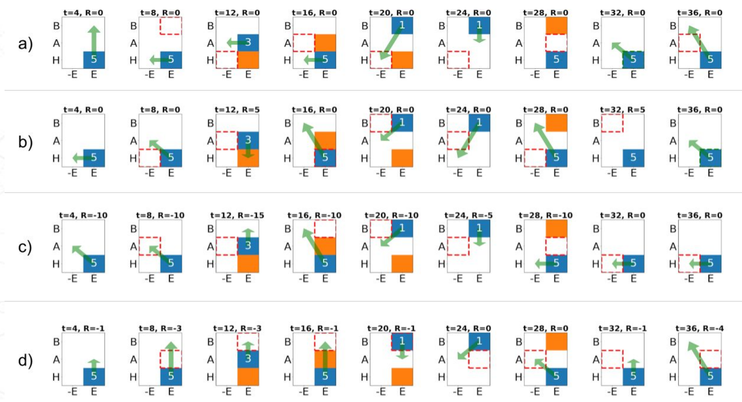
Input: human states, contextual data and set of nudges
Output: Nudges that maximize cumulative reward

For all t = 1,....,T do
  Receive context x_t
  Sample θ_t from the posterior $P(\theta|H_{t-1})$
  Select $a_t = \text{argmax}_a E(r|x_t, a, \theta_t)$
  Recommend nudge a_t and observe reward r_t
  Update $H_t = H_{t-1} \cup (x_t, a_t, r_t)$

Summary

Recommendation of nudges can be synchronized with human needs by using reinforcement learning policies. Unlike the Q-learning algorithm, the contextual bandit algorithm applies dynamic contextual data to identify the best nudge while exploring alternative nudges that are strong performers. Contextual data, when corroborated by the human in real time, may build transparency and trust.

Results



Comparison of policies that were learned by an agent when untrained, and when given three different reward functions. Nudges in the baseline (a) appeared to be random and independent of observations. The state-based reward function (b) shows a different set of nudges, which yielded low rewards on predictions. The distance-based reward function (c) depicted a more constrained set of nudges, which were closely related to the current observation. The combo-based reward function (d) demonstrated that it is possible to combine benefits from multiple reward signals.

Discussion

Because recommendations could be made frequently, we propose to limit the presentation of nudges until (1) after a stabilized period of the flow state, (2) after dwelling in a distracted state for a period of time, and (3) after task completion. It is important to integrate interest, enjoyment and concentration into flow state. Contextual data and self-reporting capture key insights about human states and decision making criteria. Additional features from the environment provide situational cues that are essential to context.

Future Work

In the future, we plan to investigate other contextual approaches, and the formulation of combo sequences of nudges. Since the human's interests and needs change over time, it is important to incorporate summarized temporal information into case representations. Finally, during online tests with real knowledge workers, attention will be given to identifying their use of heuristics and habits.

Department of Human-Centered Design and L3Harris Institute for Assured Information