Supplementary Material: Bayesian Affect Control Theory of Self

Jesse Hoey
David R. Cheriton School of Computer Science
University of Waterloo
Waterloo, ON N2L3G1
jhoey@cs.uwaterloo.ca

Tobias Schröder
Potsdam University of Applied Sciences,
Kiepenheuerallee 5,
14469 Potsdam, Germany
post@tobiasschroeder.de

Abstract

This note describes some additional information that complements the main *BayesAct-S* model described in [2]

1 Comparing Situational and Fundamental Self Sentiments

In the paper we use the following function to measure inauthenticity:

$$\mathbf{i_a}(\mathbf{s}) = ln\left(\frac{Pr(\mathbf{s_s})}{Pr(\mathbf{s_f})}\right) \tag{1}$$

However, the original statement of the theory in [1] hypothesises that inauthenticity is a difference, from which we derived that

$$\mathbf{i_a} = \mathbf{s_s} - \mathbf{s_f}\frac{1}{1-\eta} \tag{2}$$

Where $\mathbf{i_a}$, $\mathbf{s_s}$ and $\mathbf{s_f}$ represent the accumulated inauthenticity, situational self-sentiment, and fundamental self sentiment at the current time. If we use this equation to compare $Pr(\mathbf{s_s})$ with $Pr(\mathbf{s_f})$, we end up with a convolution that gives the distribution over inauthenticity:

$$Pr(\mathbf{i_a}) = Pr(\mathbf{s_s} - \mathbf{s_f}\frac{1}{1-\eta}) = Pr(\mathbf{s_s}) * Pr(-(1-\eta)^{-1}\mathbf{s_f}) \tag{3}$$

In this case, $Pr(\mathbf{i_a})$ is a probability distribution in a three dimensional space with dimensions that correspond to EPA, but the values represented are sentiment differences, not sentiments.

However, this will create inauthenticity even when the two distributions are identical. The reason is that the probability of a difference is the probability of every possible way to subtract elements from each. For example, suppose that $Pr(\mathbf{s_f})$ is a mixture of two Gaussians at $+1$ and -1 (let's label these G_f^+ and G_f^-, respectively), and $Pr(\mathbf{s_s})$ is identical (G_s^+ and G_s^-). Then $Pr(\mathbf{i_a})$ will be a distribution with three modes, one large one at zero ($G_f^+ - G_s^+$ and $G_f^- - G_s^-$), and two smaller ones at -2 and $+2$ (corresponding to $G_f^- - G_s^+$ and $G_f^+ - G_s^-$, respectively).

Therefore, we find the calculation in Equation 1 to be more suitable for the probabilistic representation in *BayesAct-S*

2 Simulations

The video shown on the *BayesAct-S* page
`http://www.cs.uwaterloo.ca/~jhoey/research/bayesact/bayesactself/`
shows an example of the simulator being used to model the interactions of the employer/daughter identity. The simulator implements the code shown in Algorithm 1. The function `SIM-ALTER` referred to in Algorithm 1 is shown below as well in pseudocode. In the main body of the paper, we describe how the function `SIM-ALTER` can be used to compute the expected situational self-sentiments caused by interacting with a number of known or unknown interactants. This is lines 1-9 in Algorithm 1. The simulator we demonstrate in the video assumes this step has already been done (results are stored for $M = 2$ identities of "mother" and "employee", along with the "stranger" or unknown identity), and proceeds with the main loop beyond line 10.

Procedure SIM-ALTER($\mathbf{s_s}, \eta, T, \mathbf{f}_a, \mathbf{f}_c$)

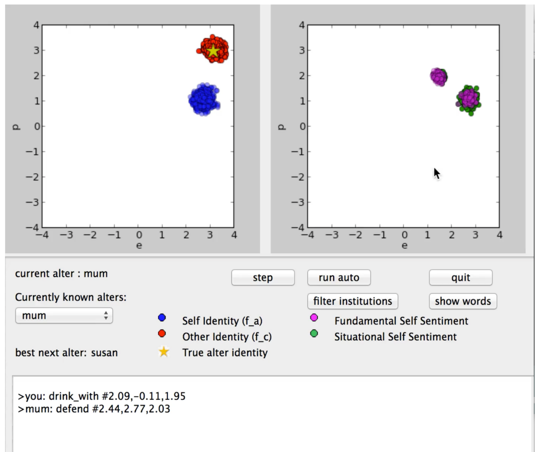
// estimate the situational self sentiments caused
// by each interactant, starting from the fundamental
for $i = 1 \dots T$ **do**
    $(\mathbf{f}_a, \mathbf{f}_c) \leftarrow$ SIMULATE($\mathbf{f}_a, \mathbf{f}_c$) // one *BayesAct* update step
    // resample $\mathbf{s_s}$ by drawing samples from $\mathbf{s_s}$ and $\mathbf{f}_a$
    // in proportions of η and $(1-\eta)$, resp.
    $\mathbf{s_s} \leftarrow \eta\mathbf{s_s} + (1-\eta)\mathbf{f}_a$
end
// we purposefully don't store the final $\mathbf{f}_a$ and $\mathbf{f}_c$
// as $\mathbf{f}_a[i], \mathbf{f}_c[i]$ here as they are only found in simulation
return $\mathbf{s_s}$

Algorithm 1: *BayesAct*-Self simulations

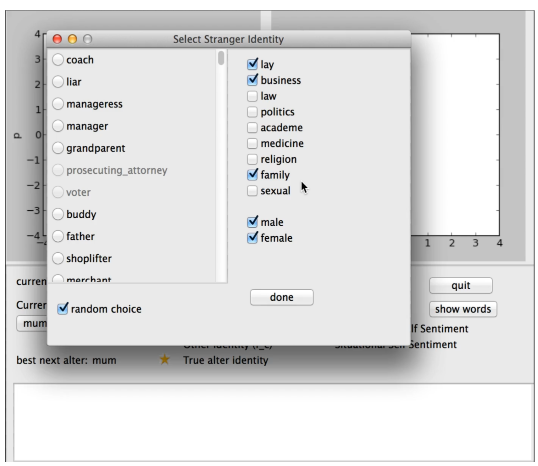
$\mathbf{s_f} \leftarrow$ set of samples representing fundamental self-sentiment
$\mathbf{s_s} \leftarrow \mathbf{s_f}$
$\mathbf{s_\circ} \leftarrow$ generic identity set for unknown interactants
$\mathbf{f}_c[j = 1 \ldots M] \leftarrow$ set of M **known** client identities (each is a set of samples)
$\mathbf{f}_c[M+1] \leftarrow \mathbf{s_\circ}$ `// an unknown identity alter (a stranger)`
$\mathbf{f}_a[j = 1 \ldots M+1] \leftarrow \mathbf{s_f}$ `// same for all interactants`
for $j = 1 \ldots M+1$ **do**
    $\mathbf{s_s}^\dagger[j] \leftarrow$ SIM-ALTER$(\mathbf{s_f}, \eta, T, \mathbf{f}_a[j], \mathbf{f}_c[j])$
end
while *True* **do**
    `// Select new interactant: the one that is predicted to`
    `// bring` $\mathbf{s_s}$ `closest to` $\mathbf{s_f}$
    for $j = 1 \ldots M+1$ **do**
        `// Combination of current self-sentiment and stored`
        `// self-sentiment expected from interacting with` j
        `// Start from` $\mathbf{s_s}^\dagger[j]$ `based on fundamentals`
        $\mathbf{s_s}' \leftarrow \eta_s \mathbf{s_s}^\dagger[j] + (1-\eta_s)\mathbf{s_s}$ `// what is` η_s`?`
        `// Possibly do some additional simulation`
        $\mathbf{s_s}' \leftarrow$ SIM-ALTER$(\mathbf{s_s}', \eta, T_s, \mathbf{f}_a[j], \mathbf{f}_c[j])$
        `// compare` $\mathbf{s_s}'$ `and` $\mathbf{s_f}$
        $\mathbf{i_a}[j] \leftarrow$ CompareBeliefs$(\mathbf{s_s}', \mathbf{s_f})$ `// KL divergence`
    end
    $j^* \leftarrow \arg\min(\mathbf{i_a}[j])$
    `// If we interact with a stranger, add a new stranger id`
    if $j^* = M+1$ **then**
        $M \leftarrow M+1$
        $\mathbf{f}_a[M+1] \leftarrow \mathbf{s_f}$
        $\mathbf{f}_c[M+1] \leftarrow \mathbf{s_\circ}$
        $\mathbf{s_s}^\dagger[M+1] \leftarrow$ SIM-ALTER$(\mathbf{s_f}, \eta, T, \mathbf{f}_a[M+1], \mathbf{f}_c[M+1])$
    end
    `// One step interaction - see [3]`
    $(\mathbf{f}_a[j^*], \mathbf{f}_c[j^*]) \leftarrow$ SIMULATE$(\mathbf{f}_a[j^*], \mathbf{f}_c[j^*])$
    $\mathbf{s_s} \leftarrow \eta \mathbf{s_s} + (1-\eta)\mathbf{f}_a[j^*]$
    $\mathbf{s_s}^\dagger[j^*] \leftarrow \eta \mathbf{s_s}^\dagger[j^*] + (1-\eta)\mathbf{s_s}$
end

Here we give a brief description of the key parts of the video.

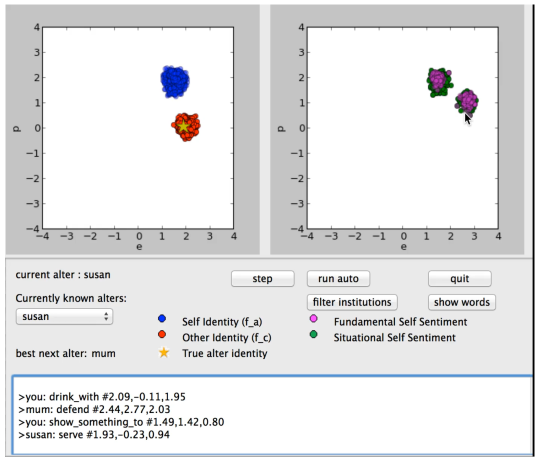
1. Start of simulation, after interaction with mother. The plot on the right shows the current fundamental self-sentiment (pink samples) and situational self-sentiment (green samples). The Evaluation (e) and Potency (p) axes are shown only, even though the simulation also uses the Activity (a) axis. The plot on the left shows the agent's estimate of her own currently enacted identity (blue samples, close to that of "daughter"), and the agent's estimate of the mother's identity (red samples, close to the true identity of "mother" as shown by the gold star). At the bottom is shown the actions taken by each agent. The actions taken are EPA vectors, and the labels are the ones closest to that EPA vector in the ACT lexicon.



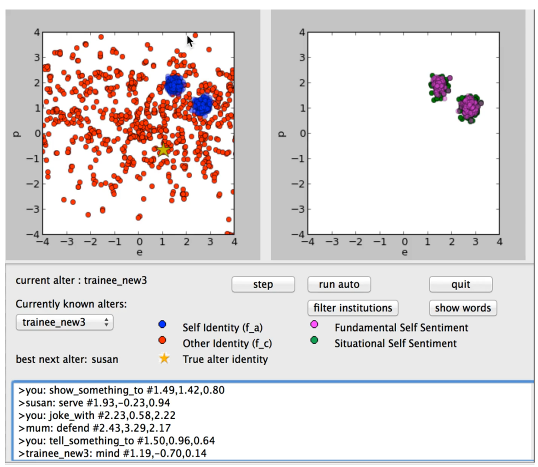
2. We select a set of institutions that are relevant to this interaction. These will constrain the set of identities that are possible for strangers (see below) and will constrain the set of behaviour labels that can be shown in the bottom text window.



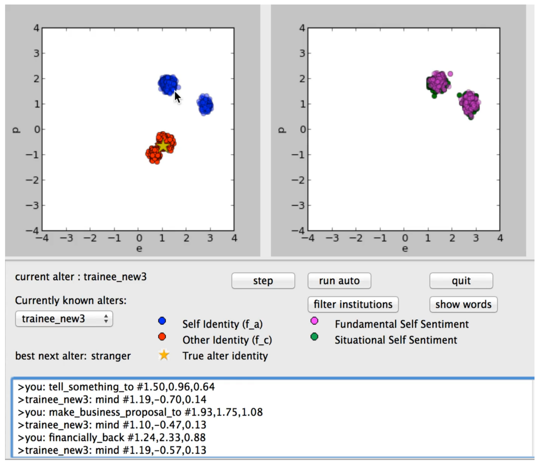
3. The next interactant is "susan" (the employee), and the enacted identity of "employer" allows the agent's situational self-sentiment to be more closely aligned with her fundamental self-sentiment.



4. Now the agent chooses to interact with a stranger. The agent's estimate of the client's identity is therefore dispersed over the space (red dots in the left plot), and the agent's estimate of her own identity (blue dots) remains bimodal because she doesn't know what identity to enact with a stranger. The agent has already done a simulation with an unknown identity to estimate the resulting situational self-sentiment.



5. After a few interactions with the stranger, the agent has learned the stranger's identity (actual value shown with the gold star).



6. The video finally shows how the simulator can be put in "auto" mode and simulate for longer periods. Each time a stranger is interacted with, its identity is randomly selected, the agent learns this identity, and a new stranger is added to the list of possibilites.

3 Bibliography

[1] David R. Heise. *Expressive Order: Confirming Sentiments in Social Actions*. Springer, 2007.

[2] Jesse Hoey and Tobias Schröder. Bayesian affect control theory of self. In *Proceedings of the AAAI Conference on Artificial Intelligence*, 2015.

[3] Jesse Hoey, Tobias Schröder, and Areej Alhothali. Bayesian affect control theory. In *Proc. of the 2013 Humaine Association Conference on Affective Computing and Intelligent Interaction (ACII 2013)*, 2013.