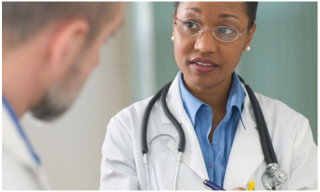
Updating Probabilistic Epistemic States in Persuasion Dialogues

Anthony Hunter[1] and Nico Potyka[2]

[1]Department of Computer Science, University College London, UK.
[2]Institute of Cognitive Science, University of Osnabrück, Germany

July 11, 2017

Bringing argumentation into persuasion technology

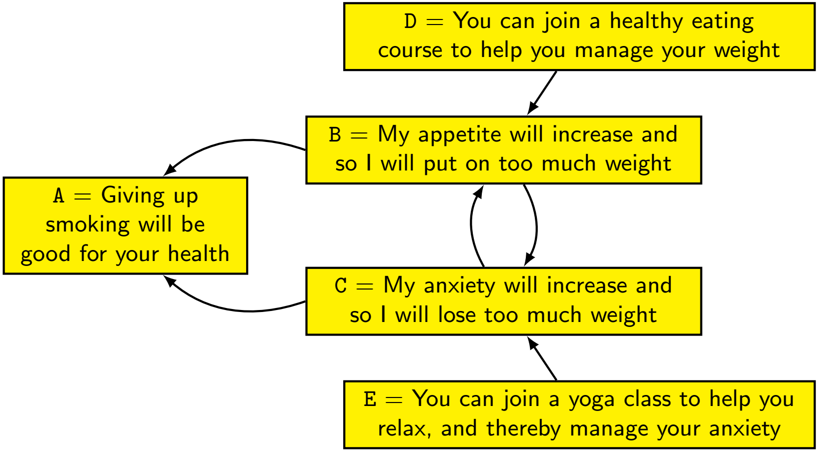


- Persuasion is an activity that involves one party trying to induce another party to believe something or to do something.
- Persuasion technologies are being developed to help people make positive changes to their behaviour (e.g. healthcare and healthy life styles).
- Current persuasion technologies either help users to explore their issues (e.g. game playing) or help users once they are persuaded to do something (e.g. diaries for recording calorie intake for weight management).
- We are aiming to bring computational models of argument into persuasion technologies.

Computational persuasion in behaviour change

Our requirements for computational persuasion via an app

1. Need asymmetric dialogues without natural language interface.
2. Need short dialogues to keep users engaged
3. Need well-chosen arguments to maximize impact
4. Need to model the user in order to be able to optimize the dialogue
5. Need to learn from previous interactions with the agent or similar agents
6. Need to model the domain to generate arguments/counterarguments



Example of argument graph for persuasion. It contains the arguments known (but not necessarily believed) by the system (i.e. the app).

Probabilistic user models

Epistemic probabilities (see Hunter and Thimm ECAI'14)

- A **mass distribution** P over $\wp(\text{Args}(G))$ s.t. $\sum_{X \subseteq \text{Args}(G)} P(X) = 1$.
- The **probability of an argument** A is $P(A) = \sum_{X \subseteq \text{Args}(G) \text{ s.t. } A \in X} P(X)$.

Rational mass distribution

P is **rational** for G iff $\forall (A, B) \in \text{Attacks}(G)$, if $P(A) > 0.5$, then $P(B) \leq 0.5$.

Example

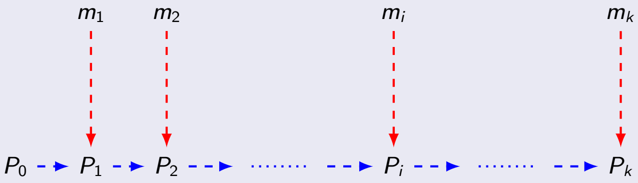
$$A \leftarrow B \leftarrow C$$

$P(A)$	$P(B)$	$P(C)$	Rational
0.6	0.9	0.9	No
0.6	0.3	0.9	Yes
0	1	0.3	Yes

Updating user models

Schematic representation of a dialogue and the user model

Let $D = [m_1, \ldots, m_k]$ be a dialogue



Each user model P_i is obtained from P_{i-1} and m_i using an update method.

Aim of dialogue w.r.t. persuasion goal A

Maximize $P_k(A)$ (i.e. according to user model, the user believes A)

Updating user models

Each move may require an update to user model

- For example, suppose an argument A is posited, then we might want to update the user model so that $P(A) = 1$.

In general, we may update by a Boolean combination of arguments.

Plan for this paper

- Some general properties for update functions
- Two specific classes of update function
  - Refinement-based update functions
  - Distance-based update functions

Properties of update functions

An update function

An **update function** as a function $U : \mathcal{P} \times \mathrm{Form} \to 2^{\mathcal{P}}$.

General properties

- **Uniqueness:** $|U(P, F)| \leq 1$.
- **Completeness:** If $F \not\equiv \bot$ then $|U(P, F)| \geq 1$.
- **Tautology:** $U(P, \top) = \{P\}$.
- **Contradiction:** $U(P, \bot) = \emptyset$.
- **Representation Invariance:** If $F \equiv G$ then $U(P, F) = U(P, G)$.
- **Idempotence:** If $U(P, F) = \{P^*\}$ then $U(P^*, F) = \{P^*\}$.
- **Order Invariance:** $U(U(P, F_1), F_2) = U(U(P, F_2), F_1)$.

Example of a satisfaction condition

- **STRICT:** P satisfies F *strictly* iff $P(F) = 1$.

Refinement-based update functions

Definition

Let L be a literal, let P be a probability distribution, and let $\lambda \in [0, 1]$.

The **refinement function** $H_\lambda : \mathcal{P} \times \{A, \neg A \mid A \in \mathrm{Args}\} \to \mathcal{P}$ is defined by $H_\lambda(P, L) = P^*$ as follows where $X \subseteq \mathrm{Args}$

$$P^*(X) = \begin{cases} P(X) + \lambda \cdot P(h_L(X)) & \text{if } X \models L \\ (1 - \lambda) \cdot P(X) & \text{if } X \models \neg L, \end{cases}$$

where $h_L(X) = X \setminus \{A\}$ if $L = A$ and $h_L(X) = X \cup \{A\}$ if $L = \neg A$ for $A \in \mathrm{Args}$.

Example

AB	P	$H_1(P, A)$	$H_1(P, \neg A)$	$H_{0.75}(P, A)$	$H_1(P, B)$
11	0.6	0.7	0.0	0.675	0.8
10	0.2	0.3	0.0	0.275	0.0
01	0.1	0.0	0.7	0.025	0.2
00	0.1	0.0	0.3	0.025	0.0

Refinement-based update functions

The **naive update function** $U_{\mathrm{na}} : \mathcal{P} \times \{A, \neg A \mid A \in \mathrm{Args}\} \to \mathcal{P}$ is

$$U_{\mathrm{na}}(P, L) = H_1(P, L)$$

Example

For dialogue $[A, B]$, the naive method gives the following updates.

	111	110	101	100	011	010	001	000
P_0					0.3	0.2	0.3	0.2
P_1	0.3	0.2	0.3	0.2				
P_2	0.6	0.4						

Note, P_2 is not rational.

Refinement-based update functions

The **trusting update function** $U_{\mathrm{tr}} : \mathcal{P} \times \mathrm{Args} \to \mathcal{P}$ is

$$U_{\mathrm{tr}}(P, A) = H_1(P, \Phi)$$

where $\Phi = \{A\} \cup \{\neg C \mid (A, C) \in \mathsf{Attacks}(G) \text{ or } (C, A) \in \mathsf{Attacks}(G)\}$.

Example

For dialogue $[A]$, the trusting method gives the following update.

	111	110	101	100	011	010	001	000
P_0					0.3	0.2	0.3	0.2
P_1			0.6	0.4				

Refinement-based update functions

The **strict update function** is a function $U_{\text{st}} : \mathcal{P} \times \mathrm{Args} \to \mathcal{P}$.

$$U_{\text{st}}(P, A) = \begin{cases} H_1(P, \Phi) & \text{if } \textit{forall}(B, A) \in \mathrm{Attacks},\ P(B) \leq 0.5 \\ P & \text{else} \end{cases}$$

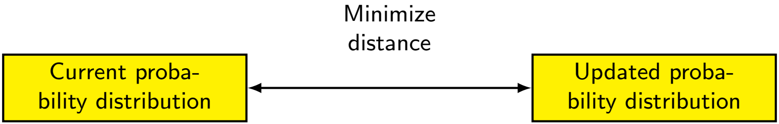
where $\Phi = \{A\} \cup \{\neg C \mid (A, C) \in \mathrm{Attacks}\}$.

Example

For dialogue $[A, C, A]$, the trusting method gives the following update.

	111	110	101	100	011	010	001	000
P_0	0.2	0.3			0.3	0.2		
P_1	0.2	0.3			0.3	0.2		
P_2			0.5				0.5	
P_3			1.0					

Distance-based update functions



- Update the epistemic state such that we minimally change the prior state.
- We can consider different change functions over $\mathcal{P}$.
    - For example, Manhattan Distance, Least Squares Distance, Maximum Distance, or KL-divergence.
    - Also, we can define distance to mimic aspects of refinement-based update.

Distance-based update functions

Some constraints on distributions (Hunter and Thimm ECAI 2014)

- **RAT:** P is *rational* iff for all $(A, B) \in \mathrm{Attacks}$, $P(A) > 0.5$ implies $P(B) \leq 0.5$.
- **COH:** P is *coherent* iff for all $(A, B) \in \mathrm{Attacks}$, $P(A) \leq 1 - P(B)$.
- **SFOU:** P is *semi-founded* iff $\mathsf{Attackers}(A) = \emptyset$ implies $P(A) \geq 0.5$.
- **FOU:** P is *founded* iff $\mathsf{Attackers}(A) = \emptyset$ implies $P(A) = 1$.
- **SOPT:** P is *semi-optimistic* iff $\mathsf{Attackers}(A) \neq \emptyset$ implies $P(A) \geq 1 - \sum_{B \in \mathsf{Attackers}(A)} P(B)$.
- **OPT:** P is *optimistic* iff $P(A) \geq 1 - \sum_{B \in \mathsf{Attackers}(A)} P(B)$.

Distance-based update functions

Definition

An **R-S-d Update Function** $U_{R,S,d} : \mathcal{P} \times \mathrm{Form} \to 2^{\mathcal{P}}$ is defined by

$$U_{R,S,d}(P, F) = \arg \min_{P' \in \mathrm{Mod}_{R,S}(F)} d(P, P').$$

where

$$\mathrm{Mod}_{R,S}(F) = \{P \in \mathcal{P} \mid P \models R, P \models_S F\}$$

and

- $R \subseteq \{RAT, COH, SFOU, FOU, SOPT, OPT\}$,
- $S \in \{STRICT, \epsilon\text{-}WEAK\}$ and
- $d \in \{d_1, d_2, d_\infty, d^X_{\mathrm{At}}, d^X_{\mathrm{Jo}}\}$.

Distance-based update functions

$$A \leftarrow B \leftarrow C$$

A	B	C	P	$U_{R_1,S,d}(P,A)$	$U_{R_1,S,d}(P,B)$	$U_{R_2,S,d}(P,A)$	$U_{R_2,S,d}(P,\top)$
0	0	0	0.2	0	0	0	0.17
0	1	0	0.5	0	1	0	0.49
0	0	1	0	0	0	0	0
0	1	1	0.1	0	0	0	0.07
1	0	0	0	0.45	0	0	0.02
1	1	0	0	0	0	0	0.5
1	0	1	0.1	0.55	0	1	0.09
1	1	1	0.1	0	0	0	0.12

Table: Illustration with $R_1 = \{COH\}$, $R_2 = \{COH, SOPT\}$, $S = STRICT$, and $d = d_2$.

Update functions

Some results

- The naive, trusting and strict refinement-based functions satisfy the general properties for literal updates.
- The trusting and strict update functions satisfy the rational and coherence properties.
- Distance-based functions using least-square distance satisfy all the general properties.
- Distance-based functions subsume the naive, trusting and strict update functions.
- There are refinement-based functions that we haven't yet captured as distance-based functions.

Discussion

Contributions

- General properties for updating probabilistic user model
- Two classes of update function
  - refinement-based
  - distance-based
- Results on general properties and inter-relationships between specific types of update function

Future work

- Methods for updating for other types of dialogue move.
- Methods for learning update functions from data.
- Evaluation with users.