A Causal Theory of Speech Acts

Chiaki Sakama
Wakayama University, Japan

LORI-VI, Sapporo, September 2017,

Background

- An **assertive speech act** commits a speaker to the truth of the expressed proposition, but she does not always have complete knowledge about the world.
- It may happen that a speaker utters a believed-true sentence which is actually false. In this case, a speaker acts truthfully but a hearer would consider the speaker untrustful.
- Whether a speech act is **truthful or not** depends on the belief state of a speaker, while whether a speech act is **trustful or not** is judged by the truth of information conveyed by the utterance.

Contribution

- We formulate assertive speech acts by agents having incomplete knowledge using a **nonmonotonic causal logic**.
- It distinguishes (un)truthful and/or (un)trustful speech acts and represents performative effects on hearers.
- A causal theory is implemented by **logic programming**.

Causal Logic

(Giunchiglia, *et al*., *Artif. Intell*. 153, 2004)

- A **causal theory** T consists of **causal rules** of the form:

  $\varphi \Rightarrow \psi$ "ψ is caused if φ is true"

  where φ and ψ are propositional formulas.

- In particular, a **constraint** is represented as

  $\varphi \Rightarrow \bot$ "φ cannot be true"

- An interpretation I (complete & consistent finite set of literals) is a **model** of T iff $I=\{ L \mid T^I \models L \}$ where $T^I = \{ \psi \mid (\varphi \Rightarrow \psi) \in T$ for some φ and $I \models \varphi \}$.

- I is a model of T iff I is the unique model of T^I .

Actions & Fluents

- $Utter_t(x,\sigma)$: an agent x utters a sentence σ at time t
- $Hold_t(\sigma)$: a sentence σ is true at time t
- $Bel_t(x,\sigma)$: an agent x believes a sentence σ at time t
- The following relations are defined:

  $Hold_t(\top) \equiv Bel_t(x,\top) \equiv \top$, $Hold_t(\bot) \equiv Bel_t(x,\bot) \equiv \bot$

  $Hold_t(\neg\sigma) \equiv \neg\, Hold_t(\sigma)$ for any x, σ, t

  where $\top$ represents true and $\bot$ represents false.

Causal Theory of Speech Acts

- A **causal theory of speech acts** $CT_{x\sigma}{}^{t}$ consists of rules:

$(\neg)Utter_t(x,\sigma) \Rightarrow (\neg)Utter_t(x,\sigma)$ (**action rule**)

``if an action (*Utter*) occurs at *t*, there is a cause of this"

$(\neg)Hold_t(\sigma) \Rightarrow (\neg)Hold_t(\sigma)$

$(\neg)Bel_t(x,\sigma) \Rightarrow (\neg)Bel_t(x,\sigma)$ (**fluent rules**)

``if a fluent (*Hold* or *Bel*) holds at *t*, there is a cause of this"

$(\neg)Hold_t(\sigma) \wedge (\neg)Hold_{t+1}(\sigma) \Rightarrow (\neg)Hold_{t+1}(\sigma)$

$(\neg)Bel_t(x,\sigma) \wedge (\neg)Bel_{t+1}(x,\sigma) \Rightarrow (\neg)Bel_{t+1}(x,\sigma)$

(**inertia rules**)

``if the truth value of a fluent (*Hold* or *Bel*) at *t* is identical with the value at *t*+1, then the value at *t*+1 is caused by persistence"

(Un)trustful/(Un)truthful Speech Acts

- A (**un**)**trustful speech act** of a sentence σ by an agent a at time t is defined as:

  Trustful$(a,\sigma,t) := CT_{a\sigma}^{t} \cup \{ Utter_t(a,\sigma) \wedge \neg Hold_t(\sigma) \Rightarrow \bot \}$

  Untrustful$(a,\sigma,t) := CT_{a\sigma}^{t} \cup \{ Utter_t(a,\sigma) \wedge Hold_t(\sigma) \Rightarrow \bot \}$

- A (**un**)**truthful speech act** of a sentence σ by an agent a at time t is defined as:

  Truthful$(a,\sigma,t) := CT_{a\sigma}^{t} \cup \{ Utter_t(a,\sigma) \wedge \neg Bel_t(a,\sigma) \Rightarrow \bot \}$

  Untruthful$(a,\sigma,t) := CT_{a\sigma}^{t} \cup \{ Utter_t(a,\sigma) \wedge Bel_t(a,\sigma) \Rightarrow \bot \}$

Properties

- **Trustful**$(a,\sigma,t) \wedge$ **Untrustful**$(a,\sigma,t) \supset \neg \mathit{Utter}_t(a,\sigma)$
- **Truthful**$(a,\sigma,t) \wedge$ **Untruthful**$(a,\sigma,t) \supset \neg \mathit{Utter}_t(a,\sigma)$
- **Trustful**$(a,\sigma,t) \wedge$ **Truthful**$(a,\sigma,t) \supset$
  $(\mathit{Utter}_t(a,\sigma) \supset \mathit{Hold}_t(\sigma) \wedge \mathit{Bel}_t(a,\sigma))$
- **Trustful**$(a,\sigma,t) \wedge$ **Untruthful**$(a,\sigma,t) \supset$
  $(\mathit{Utter}_t(a,\sigma) \supset \mathit{Hold}_t(\sigma) \wedge \neg \mathit{Bel}_t(a,\sigma))$
- **Untrustful**$(a,\sigma,t) \wedge$ **Truthful**$(a,\sigma,t) \supset$
  $(\mathit{Utter}_t(a,\sigma) \supset \neg \mathit{Hold}_t(\sigma) \wedge \mathit{Bel}_t(a,\sigma))$
- **Untrustful**$(a,\sigma,t) \wedge$ **Untruthful**$(a,\sigma,t) \supset$
  $(\mathit{Utter}_t(a,\sigma) \supset \neg \mathit{Hold}_t(\sigma) \wedge \neg \mathit{Bel}_t(a,\sigma))$

Effect of Speech Acts on Hearers

- Suppose that a speaker *a* utters a sentence σ at time *t*, which brings about a hearer *b*'s believing σ at time *t*+1. It is represented by the causal rule:

$$Utter_t(a,\sigma) \Rightarrow Bel_{t+1}(b,\sigma)$$

- A hearer would believe an utterance only when it is consistent with her own belief. The situation is represented by the constraint:

$$Bel_s(b,\neg\sigma) \wedge Bel_s(b,\sigma) \Rightarrow \bot \quad \text{for } s=t,\ t+1$$

(Mis)inform/(In)sincere

- Let *a* and *b* two agents. Define

  Inform(a,b,σ,t):= **Trustful**$(a,\sigma,t) \cup \Delta_{ab\sigma}{}^{t}$

  Misinform(a,b,σ,t):= **Untrustful**$(a,\sigma,t) \cup \Delta_{ab\sigma}{}^{t}$

  Sincere(a,b,σ,t):= **Truthful**$(a,\sigma,t) \cup \Delta_{ab\sigma}{}^{t}$

  Insincere(a,b,σ,t):= **Untruthful**$(a,\sigma,t) \cup \Delta_{ab\sigma}{}^{t}$

  where $\Delta_{ab\sigma}{}^{t}$ consists of rules:

  $(\neg)Bel_t(b,\delta) \Rightarrow (\neg)Bel_t(b,\delta)$ where $\delta \in \{\sigma, \neg\sigma\}$

  $(\neg)Bel_t(b,\delta) \wedge (\neg)Bel_{t+1}(b,\delta) \Rightarrow (\neg)Bel_{t+1}(b,\delta)$

  $Utter_t(a,\sigma) \Rightarrow Bel_{t+1}(b,\sigma)$

  $Bel_s(b,\neg\sigma) \wedge Bel_s(b,\sigma) \Rightarrow \perp$ for $s=t, t+1$

Misleading/Deceiving

- **Misinform** has the model representing **misleading**:

$\{ Utter_t(a,\sigma), Bel_t(a,\sigma), \neg Bel_t(b,\sigma), \neg Bel_t(b,\neg\sigma), \neg Hold_t(\sigma),$
$Bel_{t+1}(a,\sigma), Bel_{t+1}(b,\sigma), \neg Bel_{t+1}(b,\neg\sigma), \neg Hold_{t+1}(\sigma)\}$

``a speaker *a* utters a **believed-true** sentence σ that is **actually false**, and it causes a hearer *b*'s **acquiring the false belief**''

- **Insincere** has the model representing **deceiving**:

$\{ Utter_t(a,\sigma), \neg Bel_t(a,\sigma), \neg Bel_t(b,\sigma), \neg Bel_t(b,\neg\sigma), \neg Hold_t(\sigma),$
$\neg Bel_{t+1}(a,\sigma), Bel_{t+1}(b,\sigma), \neg Bel_{t+1}(b,\neg\sigma), \neg Hold_{t+1}(\sigma)\}$

``a speaker *a* utters a **disbelieved** sentence σ that is **actually false**, and it causes a hearer *b*'s **acquiring the false belief**''

Encoding in Logic Programming

- A causal theory of speech acts $CT_{x\sigma}^{\ t}$ is encoded into a logic program $\Pi_{x\sigma}^{\ t}$ such that

$Utter_t(x,\sigma) \leftarrow \text{not } \neg Utter_t(x,\sigma)$, $\quad \neg Utter_t(x,\sigma) \leftarrow \text{not } Utter_t(x,\sigma)$

$Hold_t(\sigma) \leftarrow \text{not } \neg Hold_t(\sigma)$, $\quad \neg Hold_t(\sigma) \leftarrow \text{not } Hold_t(\sigma)$

$Hold_{t+1}(\sigma) \leftarrow \text{not } \neg Hold_t(\sigma), \text{not } \neg Hold_{t+1}(\sigma)$

$\neg Hold_{t+1}(\sigma) \leftarrow \text{not } Hold_t(\sigma), \text{not } Hold_{t+1}(\sigma)$

$Bel_t(x,\sigma) \leftarrow \text{not } \neg Bel_t(x,\sigma)$, $\quad \neg Bel_t(x,\sigma) \leftarrow \text{not } Bel_t(x,\sigma)$

$Bel_{t+1}(x,\sigma) \leftarrow \text{not } \neg Bel_t(x,\sigma), \text{not } \neg Bel_{t+1}(x,\sigma)$

$\neg Bel_{t+1}(x,\sigma) \leftarrow \text{not } Bel_t(x,\sigma), \text{not } Bel_{t+1}(x,\sigma)$

- Then I is a model of $CT_{x\sigma}^{\ t}$ iff I is an answer set of $\Pi_{x\sigma}^{\ t}$.
- The effect of assertive speech acts is computed in **answer set programming**.