Monte Carlo Semantics

Robust Inference and Logical Pattern Processing Based on Integrated Deep and Shallow Semantics

Richard Bergmair

University of Cambridge Computer Laboratory
Natural Language Information Processing

Flatlands Meeting, Jun-06 2008

UNIVERSITY OF CAMBRIDGE

Why Inference?

- open-domain applications (QA, IE, IR, SUM)
- given two pieces of text T and H, find degree of logical
  - entailment $[\![\mathrm{BK} \rightarrow (\mathrm{T} \rightarrow \mathrm{H})]\!]$
  - or similarity $[\![\mathrm{BK} \rightarrow (\mathrm{T} \equiv \mathrm{H})]\!]$

Example

	(BK)	$\forall x : \mathsf{tall}(x) \equiv \mathsf{high}(x)$
	(T)	Including the 24m antenna, the Eiffel Tower is 325m high.
$\therefore$	(H)	How tall is the Eiffel Tower?

Why Inference?

- open-domain applications (QA, IE, IR, SUM)
- given two pieces of text T and H, find degree of logical
  - entailment $[\![\mathrm{BK} \rightarrow (\mathrm{T} \rightarrow \mathrm{H})]\!]$
  - or similarity $[\![\mathrm{BK} \rightarrow (\mathrm{T} \equiv \mathrm{H})]\!]$

Example

$$
\begin{array}{rl}
(\mathrm{BK}) & \forall x, y : \text{acquire}(x, y) \rightarrow \text{owns}(x, y) \\
(\mathrm{T}) & \text{Yamaha had acquired the guitar brand Ibanez,} \\
 & \quad \text{through its takeover of Hoshino Gakki Group,} \\
 & \quad \text{earlier this week.} \\
\hline
\therefore\ (\mathrm{H}) & \text{owns}(\text{Yamaha}, \text{Ibanez})
\end{array}
$$

What is *Robust* Inference?

...in an ideal world, we would have either

- **(YES)**

$$\vdash (\mathrm{BK}_1 \rightarrow (\mathrm{BK}_2 \rightarrow (\ldots \rightarrow (\mathrm{BK}_\mathrm{N} \rightarrow (\mathrm{T} \rightarrow \mathrm{H}))))),$$
$$\nvdash (\mathrm{BK}_1 \rightarrow (\mathrm{BK}_2 \rightarrow (\ldots \rightarrow (\mathrm{BK}_\mathrm{N} \rightarrow (\mathrm{T} \rightarrow \neg\mathrm{H})))));$$

- or **(NO)**

$$\nvdash (\mathrm{BK}_1 \rightarrow (\mathrm{BK}_2 \rightarrow (\ldots \rightarrow (\mathrm{BK}_\mathrm{N} \rightarrow (\mathrm{T} \rightarrow \mathrm{H}))))),$$
$$\vdash (\mathrm{BK}_1 \rightarrow (\mathrm{BK}_2 \rightarrow (\ldots \rightarrow (\mathrm{BK}_\mathrm{N} \rightarrow (\mathrm{T} \rightarrow \neg\mathrm{H}))))).$$

But if relevant knowledge is missing, say BK_1, we could have

- **(DON'T KNOW)**

$$\nvdash (\mathrm{BK}_2 \rightarrow (\ldots \rightarrow (\mathrm{BK}_\mathrm{N} \rightarrow (\mathrm{T} \rightarrow \mathrm{H})))),$$
$$\nvdash (\mathrm{BK}_2 \rightarrow (\ldots \rightarrow (\mathrm{BK}_\mathrm{N} \rightarrow (\mathrm{T} \rightarrow \neg\mathrm{H})))).$$

What is *Robust* Inference?

In the DON'T KNOW situation, where

$$\not\vdash \varphi \rightarrow \psi, \text{ while } \not\vdash \varphi \rightarrow \neg\psi,$$

we want to know whether or not

$$[\![\varphi \rightarrow \psi]\!] > [\![\varphi \rightarrow \neg\psi]\!],$$

and, more generally, we want to know, whether for two candidate entailments $\varphi_1 \rightarrow \psi_1$ and $\varphi_2 \rightarrow \psi_2$, we have

$$[\![\varphi_1 \rightarrow \psi_1]\!] > [\![\varphi_2 \rightarrow \psi_2]\!].$$

Outline

Outline

Model Theory: Classical Bivalent Logic

Definition

- Let $\Lambda = \langle \mathrm{p}_1, \mathrm{p}_2, \ldots, \mathrm{p}_N \rangle$ be a propositional language.
- Let $\mathrm{w} = [\mathrm{w}_1, \mathrm{w}_2, \ldots, \mathrm{w}_N]$ be a model.

The *truth value* $\| \cdot \|_{\mathrm{w}}^{\Lambda}$ is:

$$
\begin{aligned}
\| \perp \|_{\mathrm{w}}^{\Lambda} &= 0; \\
\|\mathrm{p}_i\|_{\mathrm{w}}^{\Lambda} &= \mathrm{w}_i \text{ for all } i; \\
\|\varphi \to \psi\|_{\mathrm{w}}^{\Lambda} &= \begin{cases} 1 & \text{if } \|\varphi\|_{\mathrm{w}}^{\Lambda} = 1 \text{ and } \|\psi\|_{\mathrm{w}}^{\Lambda} = 1, \\ 0 & \text{if } \|\varphi\|_{\mathrm{w}}^{\Lambda} = 1 \text{ and } \|\psi\|_{\mathrm{w}}^{\Lambda} = 0, \\ 1 & \text{if } \|\varphi\|_{\mathrm{w}}^{\Lambda} = 0 \text{ and } \|\psi\|_{\mathrm{w}}^{\Lambda} = 1, \\ 1 & \text{if } \|\varphi\|_{\mathrm{w}}^{\Lambda} = 0 \text{ and } \|\psi\|_{\mathrm{w}}^{\Lambda} = 0; \end{cases}
\end{aligned}
$$

for all formulae φ and ψ over Λ.

Model Theory: Satisfiability, Validity

Definition

- φ is *valid* iff $\|\varphi\|_{\mathrm{w}} = 1$ **for all** $\mathrm{w} \in \mathcal{W}$.
- φ is *satisfiable* iff $\|\varphi\|_{\mathrm{w}} = 1$ **for some** $\mathrm{w} \in \mathcal{W}$.

Definition

$$[\![\varphi]\!]_{\mathcal{W}} = \frac{1}{|\mathcal{W}|} \sum_{\mathrm{w} \in \mathcal{W}} \|\varphi\|_{\mathrm{w}}.$$

Corollary

- φ *is valid iff* $[\![\varphi]\!]_{\mathcal{W}} = 1$*.*
- φ *is satisfiable iff* $[\![\varphi]\!]_{\mathcal{W}} > 0$*.*

Outline

Bag-of-Words Inference (1)

assume strictly bivalent models;

$\Lambda = \{\text{socrates}, \text{is}, \text{a}, \text{man}, \text{every}\}, \quad |\mathcal{W}| = 2^5;$

$$\frac{(\mathrm{T}) \quad \text{socrates} \wedge \text{is} \wedge \text{a} \wedge \text{man}}{\therefore \ (\mathrm{H}) \quad \text{every} \wedge \text{man} \wedge \text{is} \wedge \text{socrates}};$$

$$\Lambda_{\mathrm{T}} = \{\text{a}\}, \qquad |\mathcal{W}_{\mathrm{T}}| = 2^1;$$

$$\Lambda_{\mathrm{O}} = \{\text{socrates}, \text{is}, \text{man}\}, \qquad |\mathcal{W}_{\mathrm{O}}| = 2^3;$$

$$\Lambda_{\mathrm{H}} = \{\text{every}\}, \qquad |\mathcal{W}_{\mathrm{H}}| = 2^1;$$

$2^1 * 2^3 * 2^1 = 2^5;$

Bag-of-Words Inference (2)

How to make this implication *false*?

- Choose the 1 out of 2^4 models from $\mathcal{W}_{\mathrm{T}} \times \mathcal{W}_{\mathrm{O}}$ which makes the antecedent true.
- Choose any of the $2^1 - 1$ models from $\mathcal{W}_{\mathrm{H}}$ which make the consequent false.

...now compute an expected value. Count zero for the $1 * (2^1 - 1) = 1$ model that makes this implication false. Count one, for the other $2^5 - 1$. Now

$$[\![\mathrm{T} \rightarrow \mathrm{H}]\!]_{\mathcal{W}} = 1 - \frac{1}{2^5} = 0.96875,$$

or, more generally,

$$[\![\mathrm{T} \rightarrow \mathrm{H}]\!]_{\mathcal{W}} = 1 - \frac{2^{|\Lambda_{\mathrm{H}}|} - 1}{2^{|\Lambda_{\mathrm{T}}| + |\Lambda_{\mathrm{H}}| + |\Lambda_{\mathrm{O}}|}}.$$

Outline

Language: Syllogistic Syntax

Let

$$\Lambda = \{x_1, x_2, x_3, y_1, y_2, y_3\};$$

$$\begin{aligned}
\text{All } X \text{ are } Y =& (x_1 \to_G y_1) \wedge (x_2 \to_G y_2) \wedge (x_3 \to_G y_3) \\
\text{Some } X \text{ are } Y =& (x_1 \wedge y_1) \vee (x_2 \wedge y_2) \vee (x_3 \wedge y_3) \\
\text{All } X \text{ are not } Y =& \neg \text{Some } X \text{ are } Y, \\
\text{Some } X \text{ are not } Y =& \neg \text{All } X \text{ are } Y,
\end{aligned}$$

where

$$\|\varphi \to_G \psi\| = \begin{cases} 1 & \text{if } \|\varphi\| \leq \|\psi\|, \\ \|\psi\| & \text{otherwise.} \end{cases}$$

Proof theory: A Modern Syllogism

$$\frac{}{\therefore \quad \text{All } X \text{ are } X} \; (S_1), \qquad \frac{\text{Some } X \text{ are } Y}{\therefore \quad \text{Some } X \text{ are } X} \; (S_2),$$

$$\frac{\begin{array}{l}\text{All } Y \text{ are } Z \\ \text{All } X \text{ are } Y\end{array}}{\therefore \quad \text{All } X \text{ are } Z} \; (S_3), \qquad \frac{\begin{array}{l}\text{All } Y \text{ are } Z \\ \text{Some } Y \text{ are } X\end{array}}{\therefore \quad \text{Some } X \text{ are } Z} \; (S_4),$$

$$\frac{\text{Some } X \text{ are } Y}{\therefore \quad \text{Some } Y \text{ are } X} \; (S_5);$$

Proof theory: “Natural Logic”

$$\frac{}{\therefore\ \text{All (red } X\text{) are } X}\ (\text{NL}_1), \qquad \frac{}{\therefore\ \text{All cats are animals}}\ (\text{NL}_2),$$

$$\frac{\text{Some } X \text{ are (red } Y\text{)}}{\therefore\ \text{Some } X \text{ are } Y}, \qquad \frac{\text{Some } X \text{ are cats}}{\therefore\ \text{Some } X \text{ are animals}},$$

$$\frac{\text{Some (red } X\text{) are } Y}{\therefore\ \text{Some } X \text{ are } Y}, \qquad \frac{\text{Some cats are } Y}{\therefore\ \text{Some animals are } Y},$$

$$\frac{\text{All } X \text{ are (red } Y\text{)}}{\therefore\ \text{All } X \text{ are } Y}, \qquad \frac{\text{All } X \text{ are cats}}{\therefore\ \text{All } X \text{ are animals}},$$

$$\frac{\text{All } X \text{ are } Y}{\therefore\ \text{All (red } X\text{) are } Y}, \qquad \frac{\text{All animals are } Y}{\therefore\ \text{All cats are } Y};$$

Natural Logic Robustness Properties

$$\frac{\text{Some } X \text{ are } Y}{\therefore \quad \text{Some } X \text{ are (red } Y\text{)}} > \frac{\text{Some } X \text{ are } Y}{\therefore \quad \text{Some } X \text{ are (big (red } Y\text{))}},$$

$$\frac{\text{Some } X \text{ are } Y}{\therefore \quad \text{Some (red } X\text{) are } Y} > \frac{\text{Some } X \text{ are } Y}{\therefore \quad \text{Some (big (red } X\text{)) are } Y},$$

$$\frac{\text{All } X \text{ are } Y}{\therefore \quad \text{All } X \text{ are (red } Y\text{)}} > \frac{\text{All } X \text{ are } Y}{\therefore \quad \text{All } X \text{ are (big (red } Y\text{))}},$$

$$\frac{\text{All (red } X\text{) are } Y}{\therefore \quad \text{All } X \text{ are } Y} > \frac{\text{All (big (red } X\text{)) are } Y}{\therefore \quad \text{All } X \text{ are } Y}.$$

Outline

Model Theory: Satisfiability, Validity, Expectation

Definition

$$[\![\varphi]\!]_{\mathcal{W}} = \frac{1}{|\mathcal{W}|} \sum_{\mathrm{w} \in \mathcal{W}} \|\varphi\|_{\mathrm{w}}.$$

How do we compute this in general?

Observation

- Draw w randomly from a uniform distribution over $\mathcal{W}$. Now $[\![\varphi]\!]$ is the probability that φ is true in w.
- If $\mathrm{W} \subseteq \mathcal{W}$ is a random sample over population $\mathcal{W}$, the sample mean $[\![\varphi]\!]_{\mathrm{W}}$ approaches the population mean $[\![\varphi]\!]_{\mathcal{W}}$ as $|\mathrm{W}|$ approaches $\mathcal{W}$.

Summary

This is work in progress, but could develop into a rich theoretic framework for robust textual inference and logical pattern processing.

- robust and practicable...
- ...in the worst case: does bag-of-words.

- justifiable from epistemology, logic, and linguistics;
- model theory enables inference via Monte Carlo method;
- proof theory is intuitive and well-understood;
  is entailed in classical logic and entails natural logic.