Monte Carlo Semantics

McPIET at RTE-4: Robust Inference and Logical Pattern Processing Based on Integrated Deep and Shallow Semantics

Richard Bergmair

University of Cambridge Computer Laboratory
Natural Language Information Processing

Text Analysis Conference, Nov-17 2008

UNIVERSITY OF CAMBRIDGE

Desiderata for a Theory of RTE

- Does it describe the **relevant aspects** of the systems we have **now**?
- Does it suggest ways of building **better** systems **in the future**?

A System for RTE

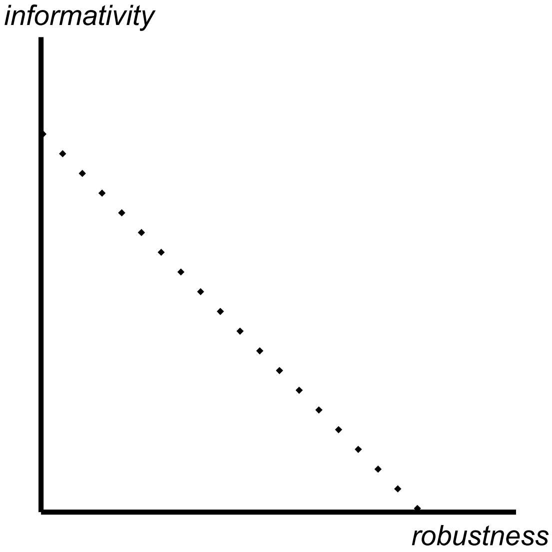
- **informativity**: Can it take into account all available relevant information?
- **robustness**: Can it proceed on reasonable assumptions, where it is missing relevant information.

Current RTE Systems

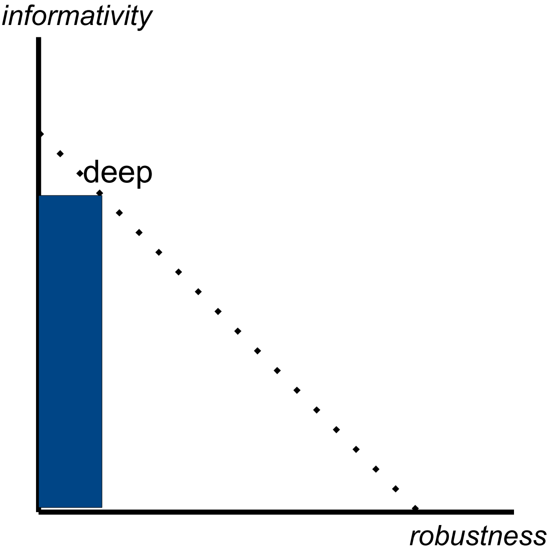
A spectrum between

- **shallow** inference
  (e.g. bag-of-words)
- **deep** inference
  (e.g. FOPC theorem proving, see Bos & Markert)

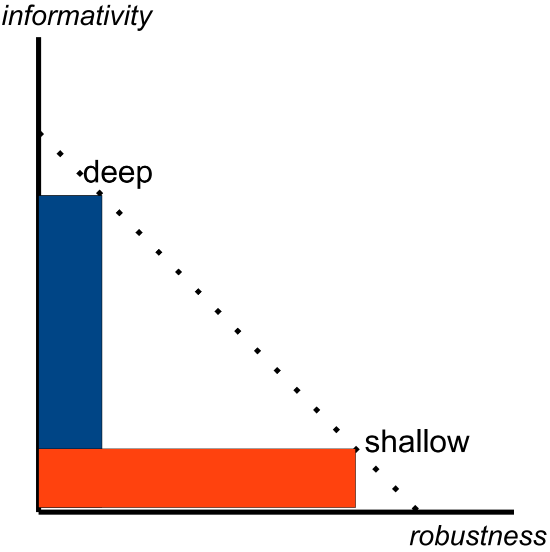
The Informativity/Robustness Tradeoff



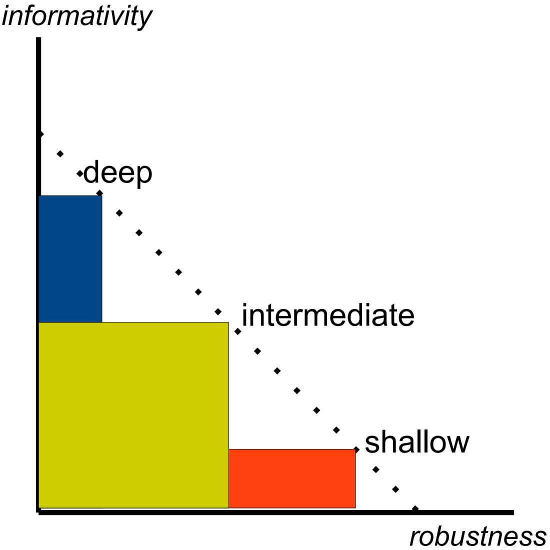
The Informativity/Robustness Tradeoff



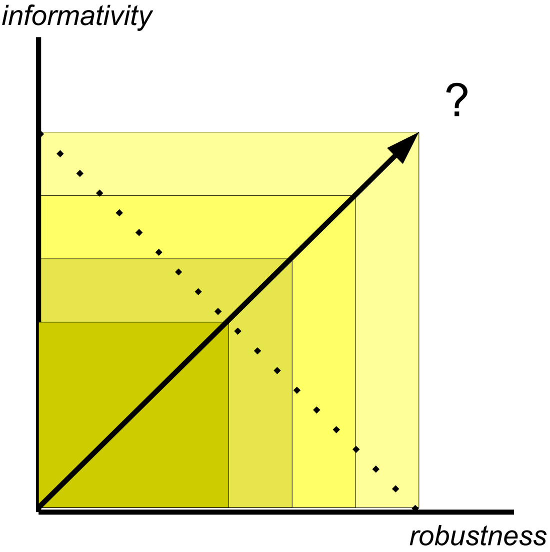
The Informativity/Robustness Tradeoff



The Informativity/Robustness Tradeoff



The Informativity/Robustness Tradeoff



Outline

Informativity, Robustness & Graded Validity

Propositional Model Theory & Graded Validity

Shallow Inference: Bag-of-Words Encoding

Deep Inference: Syllogistic Encoding

Computation via the Monte Carlo Method

Outline

Informativity, Robustness & Graded Validity

Propositional Model Theory & Graded Validity

Shallow Inference: Bag-of-Words Encoding

Deep Inference: Syllogistic Encoding

Computation via the Monte Carlo Method

Informative Inference.

predicate/argument structures

$$\top \quad > \quad \frac{\text{The cat chased the dog.}}{\rightarrow \quad \text{The dog chased the cat.}}$$

monotonicity properties, upwards entailing

$$\frac{\text{Some (grey } X\text{) are } Y}{\rightarrow \quad \text{Some } X \text{ are } Y} \quad \geq \quad \top$$

$$\top \quad > \quad \frac{\text{Some } X \text{ are } Y}{\rightarrow \quad \text{Some (grey } X\text{) are } Y}$$

Robust Inference.

monotonicity properties, upwards entailing

$$\frac{\text{Some } X \text{ are } Y}{\rightarrow \quad \text{Some (grey } X\text{) are } Y} > \frac{\text{Some } X \text{ are } Y}{\rightarrow \quad \text{Some (clean (grey } X\text{)) are } Y}$$

graded standards of proof

$$\frac{\text{Socrates is a man}}{\rightarrow \quad \text{Socrates is a man}} > \frac{\text{Socrates is a man}}{\rightarrow \quad \text{Socrates is mortal}}$$

$$\frac{\text{Socrates is a man}}{\rightarrow \quad \text{Socrates is mortal}} > \frac{\text{Socrates is a man}}{\rightarrow \quad \text{Socrates is not a man}}$$

. . . classically

(i) $T \cup \{\varphi\} \models \psi$ and $T \cup \{\varphi\} \not\models \neg\psi$;
ENTAILED / valid

(ii) $T \cup \{\varphi\} \not\models \psi$ and $T \cup \{\varphi\} \models \neg\psi$;
CONTRADICTION / unsatisfiable

(iii) $T \cup \{\varphi\} \models \psi$ and $T \cup \{\varphi\} \models \neg\psi$;
UNKNOWN / possible

(iv) $T \cup \{\varphi\} \not\models \psi$ and $T \cup \{\varphi\} \not\models \neg\psi$.
UNKNOWN / possible

. . . classically

(i) $T \cup \{\varphi\} \models \psi$ and $T \cup \{\varphi\} \not\models \neg\psi$;
ENTAILED / valid

(ii) $T \cup \{\varphi\} \not\models \psi$ and $T \cup \{\varphi\} \models \neg\psi$;
CONTRADICTION / unsatisfiable

(iii) $T \cup \{\varphi\} \models \psi$ and $T \cup \{\varphi\} \models \neg\psi$;
UNKNOWN / possible

(iv) $T \cup \{\varphi\} \not\models \psi$ and $T \cup \{\varphi\} \not\models \neg\psi$.
UNKNOWN / possible

. . . classically

(i) $T \cup \{\varphi\} \models \psi$ and $T \cup \{\varphi\} \not\models \neg\psi$;
ENTAILED / valid

(ii) $T \cup \{\varphi\} \not\models \psi$ and $T \cup \{\varphi\} \models \neg\psi$;
CONTRADICTION / unsatisfiable

(iii) $T \cup \{\varphi\} \models \psi$ and $T \cup \{\varphi\} \models \neg\psi$;
UNKNOWN / possible

(iv) $T \cup \{\varphi\} \not\models \psi$ and $T \cup \{\varphi\} \not\models \neg\psi$.
UNKNOWN / possible

. . . classically

(i) $T \cup \{\varphi\} \models \psi$ and $T \cup \{\varphi\} \not\models \neg\psi$;
ENTAILED / valid

(ii) $T \cup \{\varphi\} \not\models \psi$ and $T \cup \{\varphi\} \models \neg\psi$;
CONTRADICTION / unsatisfiable

(iii) $T \cup \{\varphi\} \models \psi$ and $T \cup \{\varphi\} \models \neg\psi$;
UNKNOWN / possible

(iv) $T \cup \{\varphi\} \not\models \psi$ and $T \cup \{\varphi\} \not\models \neg\psi$.
UNKNOWN / possible

. . . classically

(i) $T \cup \{\varphi\} \models \psi$ and $T \cup \{\varphi\} \not\models \neg\psi$;
ENTAILED / valid

(ii) $T \cup \{\varphi\} \not\models \psi$ and $T \cup \{\varphi\} \models \neg\psi$;
CONTRADICTION / unsatisfiable

(iii) $T \cup \{\varphi\} \models \psi$ and $T \cup \{\varphi\} \models \neg\psi$;
UNKNOWN / possible

(iv) $T \cup \{\varphi\} \not\models \psi$ and $T \cup \{\varphi\} \not\models \neg\psi$.
UNKNOWN / possible

. . . classically

(i) $T \cup \{\varphi\} \models \psi$ and $T \cup \{\varphi\} \not\models \neg\psi$;
ENTAILED / valid

(ii) $T \cup \{\varphi\} \not\models \psi$ and $T \cup \{\varphi\} \models \neg\psi$;
CONTRADICTION / unsatisfiable

~~(iii)~~ ~~$T \cup \{\varphi\} \models \psi$ and $T \cup \{\varphi\} \models \neg\psi$;~~
~~UNKNOWN / possible~~ (consistency)

(iv) $T \cup \{\varphi\} \not\models \psi$ and $T \cup \{\varphi\} \not\models \neg\psi$.
UNKNOWN / possible

. . . classically

(i) $T \cup \{\varphi\} \models \psi$ and $T \cup \{\varphi\} \not\models \neg\psi$;
ENTAILED / valid

(ii) $T \cup \{\varphi\} \not\models \psi$ and $T \cup \{\varphi\} \models \neg\psi$;
CONTRADICTION / unsatisfiable

~~(iii)~~ $\cancel{T \cup \{\varphi\} \models \psi}$ ~~and~~ $\cancel{T \cup \{\varphi\} \models \neg\psi}$;
~~UNKNOWN / possible~~ (consistency)

~~(iv)~~ $\cancel{T \cup \{\varphi\} \not\models \psi}$ ~~and~~ $\cancel{T \cup \{\varphi\} \not\models \neg\psi}$.
~~UNKNOWN / possible~~ (completeness)

... instead

(i) $T \cup \{\varphi\} \models_{1.0} \psi$ and $T \cup \{\varphi\} \models_{0.0} \neg\psi$;

(ii) $T \cup \{\varphi\} \models_{0.0} \psi$ and

(iii) $T \cup \{\varphi\} \models_{t} \psi$ and $T \cup \{\varphi\} \models_{t'} \neg\psi$, for $0 < t, t' < 1.0$.

   (a) $t > t'$

   (b) $t < t'$

More generally, for any two candidate entailments

- $T \cup \{\varphi_i\} \models_{t_i} \neg\psi_i$,
- $T \cup \{\varphi_j\} \models_{t_j} \neg\psi_j$,

decide whether $t_i > t_j$, or $t_i < t_j$.

...instead

(i) $T \cup \{\varphi\} \models_{1.0} \psi$ and $T \cup \{\varphi\} \models_{0.0} \neg\psi$;

(ii) $T \cup \{\varphi\} \models_{0.0} \psi$ and

(iii) $T \cup \{\varphi\} \models_{t} \psi$ and $T \cup \{\varphi\} \models_{t'} \neg\psi$, for $0 < t, t' < 1.0$.

   (a) $t > t'$
   (b) $t < t'$

More generally, for any two candidate entailments

- $T \cup \{\varphi_i\} \models_{t_i} \neg\psi_i$,
- $T \cup \{\varphi_j\} \models_{t_j} \neg\psi_j$,

decide whether $t_i > t_j$, or $t_i < t_j$.

Outline

Informativity, Robustness & Graded Validity

Propositional Model Theory & Graded Validity

Shallow Inference: Bag-of-Words Encoding

Deep Inference: Syllogistic Encoding

Computation via the Monte Carlo Method

Model Theory: Classical Bivalent Logic

Definition

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

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

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

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

Informativity, Robustness & Graded Validity

Propositional Model Theory & Graded Validity

Shallow Inference: Bag-of-Words Encoding

Deep Inference: Syllogistic Encoding

Computation via the Monte Carlo Method

Bag-of-Words Inference (1)

assume strictly bivalent valuations;
$\Lambda = \{\text{socrates}, \text{is}, \text{a}, \text{man}, \text{so}, \text{every}\}, \quad |\mathcal{W}| = 2^6;$

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

$$\begin{aligned}
\Lambda_{\mathrm{T}} &= \{\text{a}\}, & |\mathcal{W}_{\mathrm{T}}| &= 2^1; \\
\Lambda_{\mathrm{O}} &= \{\text{socrates}, \text{is}, \text{man}\}, & |\mathcal{W}_{\mathrm{O}}| &= 2^3; \\
\Lambda_{\mathrm{H}} &= \{\text{so}, \text{every}\}, & |\mathcal{W}_{\mathrm{H}}| &= 2^2;
\end{aligned}$$

$2^1 * 2^3 * 2^2 = 2^6;$

Bag-of-Words Inference (2)

How to make this implication *false*?

- Choose the 1 out of $2^4 = 16$ valuations from $\mathcal{W}_\mathrm{T} \times \mathcal{W}_\mathrm{O}$ which makes the antecedent true.
- Choose any of the $2^2 - 1 = 3$ valuations from $\mathcal{W}_\mathrm{H}$ which make the consequent false.

...now compute an expected value. Count zero for the $1 * (2^2 - 1) = 3$ valuations that make this implication false. Count one, for the other $2^6 - 3$. Now

$$[\![\mathrm{T} \rightarrow \mathrm{H}]\!]_\mathcal{W} = \frac{2^6 - 3}{2^6} = 0.95312,$$

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

Informativity, Robustness & Graded Validity

Propositional Model Theory & Graded Validity

Shallow Inference: Bag-of-Words Encoding

Deep Inference: Syllogistic Encoding

Computation via the Monte Carlo Method

Language: Syllogistic Syntax

Let

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

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

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"

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

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

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

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

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

Natural Logic Robustness Properties

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

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

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

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

Preliminary Conclusions

(a) "... you must be very naive to believe you can reason about language in logic. Even if you could, you're missing the knowledge to prove things. Even if you had that, logic would still be too computationally complex." **WRONG!**

(b) "... you must be rather ignorant to believe a machine learner will get you anywhere, if all you do is to feed it bags of words. It's just wrong from the point of view of logic, epistemology, linguistics, and whatever other theory you should care about." **WRONG!**

Preliminary Conclusions

(a) "... you must be very naive to believe you can reason about language in logic. Even if you could, you're missing the knowledge to prove things. Even if you had that, logic would still be too computationally complex." **WRONG!**

(b) "... you must be rather ignorant to believe a machine learner will get you anywhere, if all you do is to feed it bags of words. It's just wrong from the point of view of logic, epistemology, linguistics, and whatever other theory you should care about." **WRONG!**

Preliminary Conclusions

(a) "...you must be very naive to believe you can reason about language in logic. Even if you could, you're missing the knowledge to prove things. Even if you had that, logic would still be too computationally complex." **WRONG!**

(b) "...you must be rather ignorant to believe a machine learner will get you anywhere, if all you do is to feed it bags of words. It's just wrong from the point of view of logic, epistemology, linguistics, and whatever other theory you should care about." **WRONG!**

Outline

Informativity, Robustness & Graded Validity

Propositional Model Theory & Graded Validity

Shallow Inference: Bag-of-Words Encoding

Deep Inference: Syllogistic Encoding

Computation via the Monte Carlo Method

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 $\mathcal{W}$.

Outline

Informativity, Robustness & Graded Validity

Propositional Model Theory & Graded Validity

Shallow Inference: Bag-of-Words Encoding

Deep Inference: Syllogistic Encoding

Computation via the Monte Carlo Method