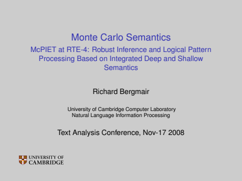


Monte Carlo Semantics

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

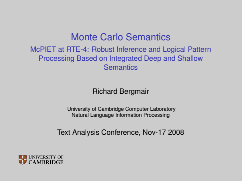
Richard Bergmair

University of Cambridge Computer Laboratory
Natural Language Information Processing

Text Analysis Conference, Nov-17 2008

UNIVERSITY OF CAMBRIDGE

- My talk is about **Monte Carlo Semantics**.
- I am currently working on this topic in Cambridge, where I have just finished the **second year** of my PhD project under the supervision of Ann Copestake.
- More particularly, the topic is robust inference and logical pattern processing based on integrated deep and shallow semantics.



Monte Carlo Semantics

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

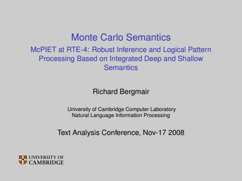
Richard Bergmair

University of Cambridge Computer Laboratory
Natural Language Information Processing

Text Analysis Conference, Nov-17 2008

UNIVERSITY OF CAMBRIDGE

- In this venue, we are seeing many talks by people who have just built a system doing textual entailment, and they come here to talk about how the system works, and what evidence they have gathered about its behaviour, experimentally.
- But, my talk has a **different format**. I will talk about my idea of robust inference, concentrating on its **theoretic foundations** within logic, combinatorics, and sampling theory.



Monte Carlo Semantics

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

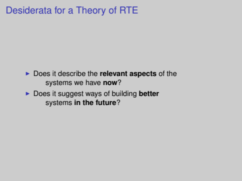
Richard Bergmair

University of Cambridge Computer Laboratory
Natural Language Information Processing

Text Analysis Conference, Nov-17 2008

UNIVERSITY OF CAMBRIDGE

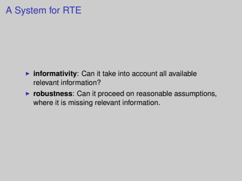
- I believe, this talk should be highly relevant to this community, nevertheless. Because, if we are building and evaluating RTE systems, then it would be useful to have a theoretic framework for thinking and talking about RTE.
- . . . and while we have seen many different **systems** for RTE, **a theoretic framework** of this kind is still largely lacking.
- In order to determine whether or not a theory of RTE is in fact useful in this sense, we have to ask ourselves two questions.



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**?

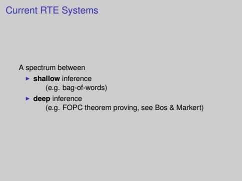
- First: Does it describe the relevant aspects of the broad range of systems that have been, and are being, built **now**.
- Second: Does it suggest possible paths of development that may lead to better systems **in the future**.
- In an attempt to answer this second question, I have started to build the McPIET Monte Carlo Pseudo Inference Engine for Text. However, I have only recently started development, and any conclusions that might be drawn on its performance, would be mere speculation at this point in time.
- In the present talk, I will concentrate on the first question. Does my theory describe the relevant aspects of the systems we currently have?



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.

- What is a good system? What are the relevant aspects?
- A good system is one that is both informative and robustness, and the relevant aspects of a system are those properties that give rise to its informativity and robustness.
- By informativity, I mean the ability of a system to take into account all available relevant information.
- By robustness, I mean the ability of a system to proceed on reasonable assumptions, even where relevant information is missing.



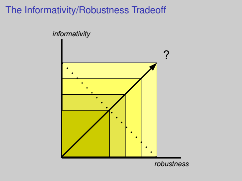
Current RTE Systems

A spectrum between

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

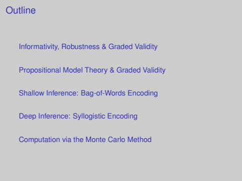
- What systems do we currently have?
- They can perhaps be situated anywhere in a spectrum between shallow and deep systems.
- For example systems making use of textual representations similar to bag-of-words techniques, might be classified as shallow.
- On the other hand, systems making use of logical representations such as FOPC, might be classified as deep.

The Informativity/Robustness Tradeoff



The Informativity/Robustness Tradeoff

- What we observe about these systems in practice is a tradeoff between robustness and informativity.
- Deep systems are very informative, yet they are not robust enough.
- Shallow systems are very robust, yet they are not informative enough.
- And most systems we see in practice, are intermediate-level systems that provide some intermediate level of both robustness and informativity.
- The goal of deep/shallow integration is to escape this tradeoff and construct a system that is both informative and robust. – And this is an interesting open problem.



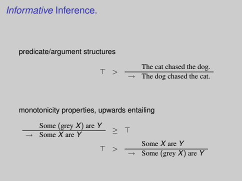
Outline

- 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

- The notion which is at the very core of my theory is that of graded validity. The first half of my talk will be dedicated to defining this notion and the theory surrounding it.
- I will first talk in some more detail about informativity and robustness. I'll give some examples, and I'll define them as theoretic notions in terms of graded validity.
- I will then go on to show how graded validity relates to the notion of validity we are used to from classical logic, and how the practical design goals of informativity and robustness relate to the theoretic notions of consistency and completeness.
- And finally, I will show how we can generalize from classical validity to graded validity within a model theoretic framework.
- Hopefully, this should give you a good idea of what my theory actually is.

- Concerning that theory, I basically make only one claim, at this point in time: That it is a useful way of thinking and talking about the robustness and informativity characteristics of current shallow and deep systems.
- I will therefore go on to establish bag-of-words techniques as a special case of robust inference within my theory, on the shallow end of the spectrum.
- And then I will similarly establish FOPC theorem proving as a special case, on the deep end of the spectrum, considering the syllogistic fragment in greater detail.
- Finally, I will then briefly mention Monte Carlo Semantics, which, I believe, is a promising way forward, although I cannot, at this point, make any strong claims concerning its actual performance.

Informative Inference.



Informative Inference.

predicate/argument structures

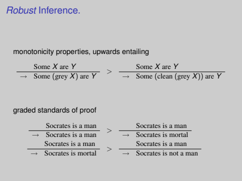
$$\top > \frac{\text{The cat chased the dog.}}{\rightarrow \text{The dog chased the cat.}}$$

monotonicity properties, upwards entailing

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

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

- Let me start by giving an example of what I mean by informativity: The cat chased the dog, therefore the dog chased the cat. This is supposed to be less than true – classically we would say: it's false.
- For the most naive bag-of-words inference system, this kind of inference would be permissible, however, since we have exactly the same words in the antecedent and the consequent.
- The important point here, is that there is information in those texts, concerning predicate/argument structure, that is not taken into account by bag-of-words comparisons. So bag-of-words inference fails on informativity here, where a theorem prover would easily get this example right.
- Generally, what one would like to do, is to make a list of example inferences, like these. An inference engine would be considered informative, if it gets all of them right.



Robust Inference.

monotonicity properties, upwards entailing

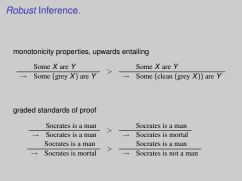
$$\frac{\text{Some } X \text{ are } Y}{\rightarrow \ \text{Some (grey } X\text{) are } Y} > \frac{\text{Some } X \text{ are } Y}{\rightarrow \ \text{Some (clean (grey } X\text{)) are } Y}$$

graded standards of proof

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

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

- We can now go on to think about robustness.
- In this first example, let's substitute elephants for X.
- We know that all elephants are grey.
- But let's assume that the machine doesn't know this. This piece of background knowledge is mentioned nowhere in the antecedent, and a background theory of common sense knowledge of this kind is probably not quite complete. So this is a very realistic scenario.
- Given that all elephants are grey, we would consider the first inference valid. If some elephants are Y, then some grey elephants are Y. Given that not all elephants are clean, we would consider the second inference invalid. It is not the case that, if some elephants are Y, then some clean grey elephants are Y.



Robust Inference.

monotonicity properties, upwards entailing

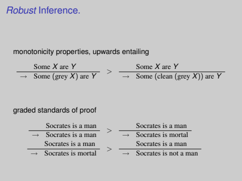
$$\frac{\text{Some } X \text{ are } Y}{\rightarrow \ \text{Some (grey } X\text{) are } Y} > \frac{\text{Some } X \text{ are } Y}{\rightarrow \ \text{Some (clean (grey } X\text{)) are } Y}$$

graded standards of proof

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

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

- A theorem prover would reach a different conclusion. In the absence of our background knowledge, it would consider them both to be equally invalid. We lose the discriminative power, to distinguish between the two cases, because we are missing relevant information.
- However, we could have proceeded on a quite reasonable assumption: We could have simply assumed that two illegal insertions of this kind are in some sense always worse than only one.
- So theorem proving fails on robustness. Curiously, the much simpler bag-of-words technique gets this right.
- It is also worth noting, that these robustness properties cannot be expressed simply as valid or unsatisfiable example inferences. Rather, they are expressed as comparisons between possible example inferences. . . . and this is why we need graded validity if we want to achieve this kind of robustness.



Robust Inference.

monotonicity properties, upwards entailing

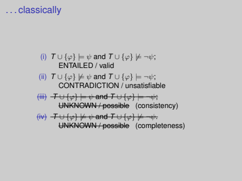
$$\frac{\text{Some } X \text{ are } Y}{\rightarrow\ \text{Some (grey } X\text{) are } Y} > \frac{\text{Some } X \text{ are } Y}{\rightarrow\ \text{Some (clean (grey } X\text{)) are } Y}$$

graded standards of proof

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

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

- I will now go on to compare the idea of graded validity to the classical notion of validity.



. . . classically

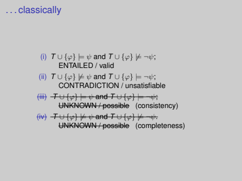
(i) $T \cup \{\varphi\} \models \psi$ and $T \cup \{\varphi\} \not\models \neg\psi$;
ENTAILED / valid

(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~~ (completeness)

- Classically, we can prove a given candidate entailment of the form $\phi \rightarrow \psi$, using the deduction theorem as follows. We start with a theory of background knowledge *T* and add to that theory the antecedent ϕ. There are now four cases.
- Case 1: If we can prove the consequent ψ, and we cannot prove its negation $\neg\psi$, then the implication is valid.
- Case 2: If we cannot prove the consequent ψ, but we can prove its negation $\neg\psi$, then the implication is unsatisfiable.
- But, combinatorically, there are two more possibilities here. Case 3: We might be able to prove both. Case 4: We might be able to prove neither.



. . . classically

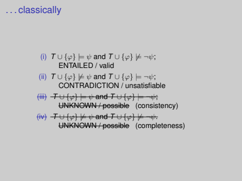
(i) $T \cup \{\varphi\} \models \psi$ and $T \cup \{\varphi\} \not\models \neg\psi$;
ENTAILED / valid

(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~~ (completeness)

- In classical logic, we would always require any theory to be consistent, so that it does not prove two contradictory theses. In other words: We require that case 3 cannot occur.
- Furthermore, we always require a theory to be complete, so that among two contradictory theses, one must always be provable. In other words: We require that case 4 cannot occur.
- Let's think about what this means in practice.
- A background theory could consist of meaning postulates such as "$\forall x : \text{cat}(x) \rightarrow \text{animal}(x)$", which could be derived from the WordNet noun hyponymy hierarchy, or "$\forall x, y, z : \text{buy-from}(x, y, z) \equiv \text{sell-to}(z, y, x)$" could be derived from a role-labelled verb lexicon. Given careful knowledge engineering we might be able to ensure that a background theory of this kind is consistent.



. . . classically

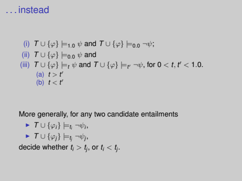
(i) $T \cup \{\varphi\} \models \psi$ and $T \cup \{\varphi\} \not\models \neg\psi$;
ENTAILED / valid

(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~~ (completeness)

- But what about completeness? Say we are trying to prove “Socrates is a man, therefore Socrates is mortal”. In the empty background theory we can prove neither that Socrates is mortal, nor that he isn’t. The background theory is incomplete.
- We would have to add knowledge to the background theory, for example saying that every man is mortal, or that no man is mortal, etc.
- Now, I think that a completeness property of the kind required by classical logic would be very nice to have, but is entirely unrealistic in practice.
- I think that this case 4, far from being nonexistent, will in practice probably be the most common case, with cases 1 and 2 occuring only as limit cases of theoretical interest.



. . . instead

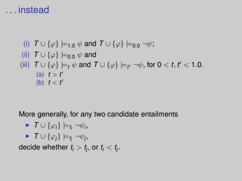
(i) $T \cup \{\varphi\} \models_{1.0} \psi$ and $T \cup \{\varphi\} \models_{0.0} \neg\psi$;
(ii) $T \cup \{\varphi\} \models_{0.0} \psi$ and
(iii) $T \cup \{\varphi\} \models_{t} \psi$ and $T \cup \{\varphi\} \models_{t'} \neg\psi$, for $0 < t, t' < 1.0$.
    (a) $t > t'$
    (b) $t < t'$

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 $t_i < t_j$.

- The solution I propose is to work with degrees of validity, rather than drawing the classical dichotomous distinction between validity and unsatisfiability.
- We would still have the classical two cases of validity, and unsatisfiability.
- If the degree of validity of ψ is 1.0, while the degree of validity of $\neg\psi$ is 0.0, then we have classical validity.
- If the degree of validity of ψ is 0.0, while the degree of validity of $\neg\psi$ is 1.0, then we have classical unsatisfiability.
- However, classical logic with its dichotomous validity notion is ignorant of this new case 3.



. . . instead

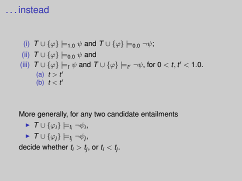
(i) $T \cup \{\varphi\} \models_{1.0} \psi$ and $T \cup \{\varphi\} \models_{0.0} \neg\psi$;
(ii) $T \cup \{\varphi\} \models_{0.0} \psi$ and
(iii) $T \cup \{\varphi\} \models_{t} \psi$ and $T \cup \{\varphi\} \models_{t'} \neg\psi$, for $0 < t, t' < 1.0$.
  (a) $t > t'$
  (b) $t < t'$

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 $t_i < t_j$.

- What we would like to do in case 3, is the following:
- We acknowledge, that neither the conclusion ψ, nor its negation $\neg\psi$ is classically provable, but we still want to know, which **is more provable** than the other.
- We acknowledge that, on the basis of the given knowledge, neither the conclusion nor its negation are supported perfectly well, but we still want to know, which of the two is **better** supported by the knowledge we have.
- Our degrees of validity are now supposed to support this kind of comparison, and allow, within case 3, a distinction into two subcases: Case (a), where ψ is more valid than $\neg\psi$; and case(b), where ψ is less valid than $\neg\psi$.



. . . instead

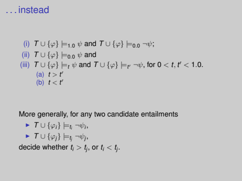
(i) $T \cup \{\varphi\} \models_{1.0} \psi$ and $T \cup \{\varphi\} \models_{0.0} \neg\psi$;
(ii) $T \cup \{\varphi\} \models_{0.0} \psi$ and
(iii) $T \cup \{\varphi\} \models_{t} \psi$ and $T \cup \{\varphi\} \models_{t'} \neg\psi$, for $0 < t, t' < 1.0$.
  (a) $t > t'$
  (b) $t < t'$

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 $t_i < t_j$.

- More generally: Given 800 candidate entailments, I do not simply want to partition them into a set of valid entailments and a set of unsatisfiable entailments. I want to order them from left to right.
- On the left end, I want to have tautologies like *Socrates is a man, therefore Socrates is a man*,
- On the right end, I want to have contradictions like *Socrates is a man, therefore Socrates is not a man*,
- And in between, I want to have various contingencies like *Socrates is a man, therefore Socrates is mortal*, ordered by their degrees of validity.



. . . 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) $t < t'$

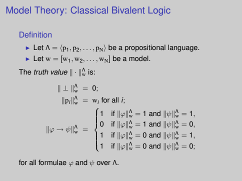
More generally, for any two candidate entailments

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

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

- Given such an ordering, and prior knowledge saying, for example, that I expect 300 yes-answers, 200 don't know answers, and 300 no-answers, I can use that ordering and determine appropriate cutoffs.
- This gives me robustness, since, instead of insisting that candidate entailments be perfectly valid or unsatisfiable logically, I now have a more lenient way of saying what exactly it means for an entailment to be good enough to be a yes-answer, or a no-answer. And everything else is then a don't know answer.

- I will now go on to define, how exactly I can determine, the degree of validity for a given candidate entaiment as a concrete number.



Model Theory: Classical Bivalent Logic

Definition

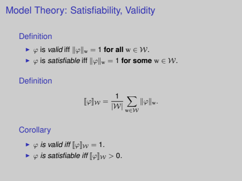
- Let $\Lambda = \langle p_1, p_2, \ldots, p_N \rangle$ be a propositional language.
- Let $w = [w_1, w_2, \ldots, w_N]$ be a model.

The *truth value* $\| \cdot \|_w^\Lambda$ is:

$$\begin{aligned} \| \perp \|_w^\Lambda &= 0; \\ \|p_i\|_w^\Lambda &= w_i \text{ for all } i; \\ \|\varphi \to \psi\|_w^\Lambda &= \begin{cases} 1 & \text{if } \|\varphi\|_w^\Lambda = 1 \text{ and } \|\psi\|_w^\Lambda = 1, \\ 0 & \text{if } \|\varphi\|_w^\Lambda = 1 \text{ and } \|\psi\|_w^\Lambda = 0, \\ 1 & \text{if } \|\varphi\|_w^\Lambda = 0 \text{ and } \|\psi\|_w^\Lambda = 1, \\ 1 & \text{if } \|\varphi\|_w^\Lambda = 0 \text{ and } \|\psi\|_w^\Lambda = 0; \end{cases} \end{aligned}$$

for all formulae φ and ψ over Λ.

- Let me start out by giving you a little refresher on how model theory works for a classical bivalent propositional logic.
- In the language of propositional logic, we have three kinds of formulae.
  - We can use the falsity constant, which always has the truth value zero.
  - We can state atomic propositions, which always get their truth value assigned, according to a valuation, to be either zero or one.
  - If phi is a formula and psi is a formula, then so is the implication “phi implies psi” and the truth value of that formula is determined according to the well known truth table.



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** $\mathrm{w} \in \mathcal{W}$.

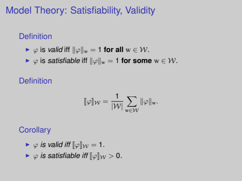
Definition

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

Corollary

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

- The model theoretic notion of validity is set up on that basis, by saying that phi is valid, i.e. phi is a tautology, iff the truth value of that formula phi is one **for all** valuations.
- That is: For every possible assignment of truth values to the atomic propositions that occur in phi, we want the truth value of the whole formulae phi to be one. Then we have a tautology. Then we have a valid formula.
- And we say that a formula is satisfiable if, **for some** valuation, the truth value of phi is one.
- That is: We want to be able to come up with some assignment of truth values to atomic propositions that make the truth value of the whole formulae phi one.



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** $\mathrm{w} \in \mathcal{W}$.

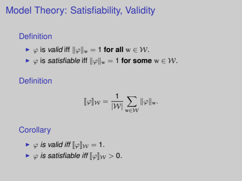
Definition

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

Corollary

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

- Now what I do instead, is to take the same setup, and use a different mode of aggregating across truth values for the different valuations.
- Validity means taking a minimum. Satisfiability means taking a maximum.
- What I'm doing instead is to take an arithmetic mean.
- This is informationally stronger than the other two, because given the arithmetic mean, we can still determine whether the formula was valid or satisfiable, but not the other way around.
- In particular: We know that the formula is valid, when the arithmetic mean is one, and we know that it is satisfiable, when it's strictly greater than zero.



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** $\mathrm{w} \in \mathcal{W}$.

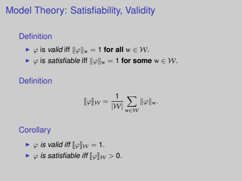
Definition

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

Corollary

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

- But the crucial thing, is that this measure can also take on values between zero and one. And this is
  - my notion of graded validity
  - the standard of proof
  - the degree to which a theory supports a conclusion



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** $\mathrm{w} \in \mathcal{W}$.

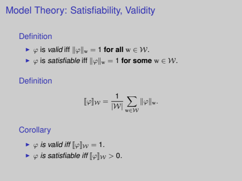
Definition

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

Corollary

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

- This definition also has a probabilistic interpretation.
- Here, one can think of $\|\chi\|$ as a random variable indicating the truth value taken on for $\|\chi\|_{w}$, when a valuation *w* is chosen from $\mathcal{W}$ at random.
- The value of $[\![\chi]\!]$ is then quite simply the probability that the truth value of χ, for such a valuation *w* chosen at random, is 1, assuming for this choice a uniform distribution.



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** $\mathrm{w} \in \mathcal{W}$.

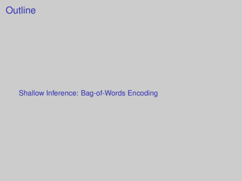
Definition

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

Corollary

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

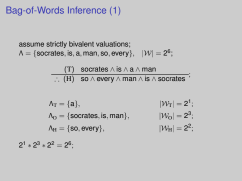
- From the point of view of traditional objectivist probability, the question arises: Why should this distribution be uniform, rather than anything else? In response to this question, one might imagine an assumption of maximum entropy, i.e. maximum uncertainty, regarding this choice of a valuation.
- From the point of view of De Finetti's subjectivist probability, this question does not even arise. The question, then, is not "Why assume a uniform distribution?", but rather "Why not?" – in the absence of any certain information contradicting such an assumption.



Outline

Shallow Inference: Bag-of-Words Encoding

- So now that we know what graded validity is, how do we use it for reasoning purposes?
- As I've said before, I'll answer this question in two steps.
- First, I will look at how this notion of graded validity can serve as a theoretical justification for bag-of-words reasoning.
- Then, I will show that we can equally apply it to deeper level reasoning.



Bag-of-Words Inference (1)

assume strictly bivalent valuations;

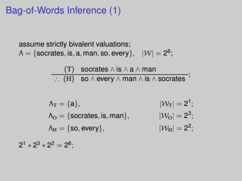
$\Lambda = \{\text{socrates}, \text{is}, \text{a}, \text{man}, \text{so}, \text{every}\}, \quad |\mathcal{W}| = 2^6;$

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

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

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

- My theoretical justification for bag-of-words reasoning is essentially based on the idea of this bag-of-words representation.
- That is: When we do bag-of-words inference, then essentially what we're doing is to treat the words in a piece of text, as if they were atomic propositions and the texts themselves, as if they were conjunctions of such atomic propositions.
- As an example, let's consider the Woody Allen mood of the syllogism. Socrates is a man, therefore every man is Socrates.



Bag-of-Words Inference (1)

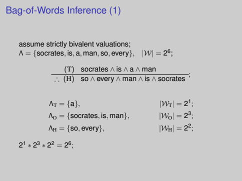
assume strictly bivalent valuations;
$\Lambda = \{\mathrm{socrates}, \mathrm{is}, \mathrm{a}, \mathrm{man}, \mathrm{so}, \mathrm{every}\}, \quad |\mathcal{W}| = 2^6;$

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

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

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

- If we write this formula down like this, then the degree of validity for that implication is simply a function of the sizes of three sets:
  - the set of words that appear only in the antecedent, but not in the consequent.
  - the set of words that appear both in the antecedent and in the consequent;
  - the set of words that appear only in the consequent, but not in the antecedent.



Bag-of-Words Inference (1)

assume strictly bivalent valuations;

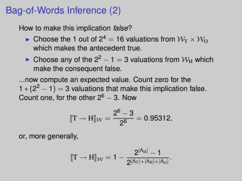
$\Lambda = \{\mathsf{socrates}, \mathsf{is}, \mathsf{a}, \mathsf{man}, \mathsf{so}, \mathsf{every}\}, \quad |\mathcal{W}| = 2^6;$

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

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

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

- Logically, we are looking at six atomic propositions, so there are $2^6 = 64$ ways of assigning truth values to these five propositions.
- There are $2^1 = 2$ ways of assigning truth values to the propositions in the antecedent.
- There are $2^3 = 8$ ways of assigning truth values to the propositions in the overlap set.
- There are $2^2 = 4$ ways of assignign truth values to the propositions in the consequent.



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.

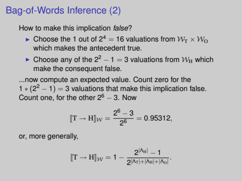
...now compute an expected value. Count zero for the $1 * (2^2 - 1) = 3$ valuations that make this implication false. Count one, for the other $2^6 - 3$. Now

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

or, more generally,

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

- Now how do we make this implication false?
- In order to make an implication false, we must make its antecedent true, and its consequent false.
- In order to make this antecedent, which is a conjunction of four propositions, true, we have to make each of its conjuncts true. So out of the 2^4 possible ways of assigning truth values to these four conjuncts, only one way makes it true, namely that assignment which makes all the conjuncts true.
- Now let's look at the consequent. Of the five words in this consequent, we have already assigned the truth value one to three of them. So there are two more proposition which need a truth value.
- Here we want to assign the value zero, in order to make the consequent false, which we can do in three ways. The only assignment we cannot choose is the one that makes them both true. The other three choices make the conjunction false.



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

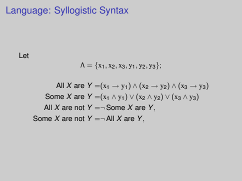
$$[\![\mathrm{T} \rightarrow \mathrm{H}]\!]_{\mathcal{W}} = \frac{2^6 - 3}{2^6} = 0.95312,$$

or, more generally,

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

- So we know that exactly three out of the 2^6 possible valuations make this implication false. Now let's count zero for these three assignments, and let's count one for all the others, and let's sum them up, and divide them by the number of valuations 2^6, and that gives us the degree of validity, which is, in our particular example this number: 0.95.
- More importantly, we can write down the degree of validity as a closed formula.
- And, not surprisingly, we find that this formula is, in fact a bag-of-words overlap measure.
- What we have shown here is that, if you take this robust inference engine, and you give it no information in the semantic representation, beyond that which is available from a simple tokeniser, then the inference mechanism will reduce to simply measuring bag-of-words overlap. It will have the same informativity and robustness properties as this well-known

- But now the question is: What happens if you make the semantic representations more informative. Can we correctly apply the information that's present in a deeper-level semantic representation?
- To answer that question, let's look at the deep end of the spectrum, where we assume that we have unambiguous information about predicate predicate-argument structures, quantifiers, and their scopes.



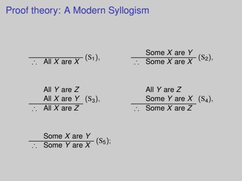
Language: Syllogistic Syntax

Let

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

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

- In this case, our method would apply a syllogistic representation.
- So rather than just looking at a text as a conjunction of words, taken as atomic propositions, . . .
- . . . , we now formulate a sentence like “All X are Y” by taking that to be a compound formula.
- We simply assume that such a formula would be a predication about three individuals 1, 2, and 3, in some artificial domain.
- So, saying that “All X are Y” would mean that “If one is an x, then 1 is a Y, AND, if 2 is an x, then 2 is a y, AND, if 3 is an x, then 3 is a Y.”
- Similarly, “Some X are Y” would mean that “One is an x and one is y, OR, two is an x and two is a y, OR, three is an x and three is a y.”



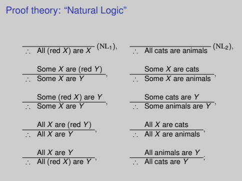
Proof theory: A Modern Syllogism

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

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

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

- If we represent these sentences in that way, it turns out that the following theorems are provable. And these theorems are the axioms of the syllogism.
- So it turns out that our logic does, in fact, have the syllogism as a fragment.



Proof theory: “Natural Logic”

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

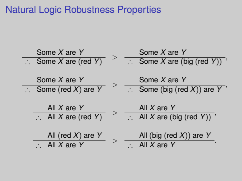
$$\frac{\text{Some } X \text{ are (red } Y\text{)}}{\therefore\quad \text{Some } X \text{ are } Y},\qquad \frac{\text{Some } X \text{ are cats}}{\therefore\quad \text{Some } X \text{ are animals}},$$

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

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

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

- Why should we care about the syllogism? The important thing is that the syllogism also proves all of the monotonicity properties that are sometimes dealt with under the name “natural logic”



Natural Logic Robustness Properties

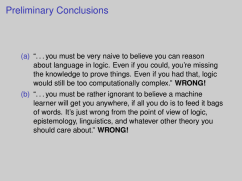
$$\frac{\text{Some } X \text{ are } Y}{\therefore \ \text{Some } X \text{ are (red } Y\text{)}} > \frac{\text{Some } X \text{ are } Y}{\therefore \ \text{Some } X \text{ are (big (red } Y\text{))}},$$

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

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

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

- So we get all the right theorems to be provable, but what about the theorems that are not perfectly provable? In this case, we still get the right robustness properties!

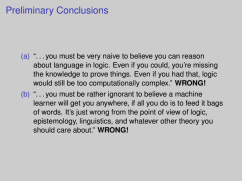


Preliminary Conclusions

(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!**

- Let me stop here, to draw some preliminary conclusions that we might draw from the theoretic results presented so far.
- Let's consider these two statements. . . .
- Although this is the sort of thing that we don't read much about these days, this separation of the world of NLP into (a) and (b), is still a deeply entrenched paradigm.
- And if I were sitting in the audience right now, I'd probably be asking myself whether what I'm hearing is an (a)-talk or a (b)-talk.
- This is why, I would like to emphasize at this point, that I subscribe neither to viewpoint (a) nor to viewpoint (b) exclusively.

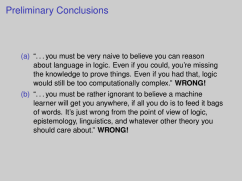


Preliminary Conclusions

(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!**

- I have talked about classical logic, subscribing to the viewpoint that the classical dichotomous notion of validity, and the associated kinds of completeness and consistency assumptions are just not practical, when it comes to reasoning within a theory containing common-sense or real-world knowledge. I believe that probability is just a better model of the epistemological phenomena we are seeing in common-sense reasoning with natural language. – this is a viewpoint we would associate with the (b)-world.

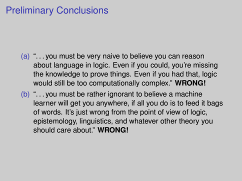


Preliminary Conclusions

(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!**

- On the other hand, I have talked about bag-of-words encodings for text, essentially claiming that if you feed your machine learner bags of words, you are forcing a semantic interpretation on your texts, which is, from a logical-semantic point of view, no less naive than saying that a text is just a conjunction, the conjuncts of which are atomic propositions symbolized by words. My viewpoint on this is that existing tools for semantic composition can do much better than that. – this is a viewpoint we would associate with the (a)-world.

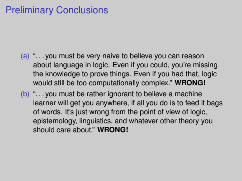


Preliminary Conclusions

(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!**

- In and of themselves, these viewpoints are quite unsurprising.
- What is new, however, about my viewpoint, is the fact that those two propositions are taken to be independent aspects of the same theory, rather than belonging to two contradictory theories.
- And, as a result, we can draw two interesting conclusions, in response to viewpoint (a) and viewpoint (b).

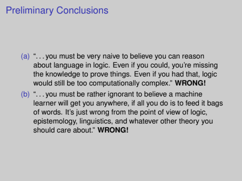


Preliminary Conclusions

(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!**

- In response to viewpoint (a), we can now say that knowledge and computational complexity are issues that are completely separate from the question of whether or not logic is a useful theoretic framework for approaching textual inference.
- It is all a question of how one represents text in logic.
- In the case of a bag-of-words representation, all the knowledge that is required is in the identities of the logical variables, and computational complexity is as little as that of evaluating a simple arithmetic expression.



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!**

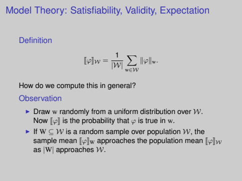
(b) "…you must be rather ignorant to believe a machine learner will get you anywhere, if all you do is to feed it bags of words. It's just wrong from the point of view of logic, epistemology, linguistics, and whatever other theory you should care about." **WRONG!**

- In response to viewpoint (b), we can now say that bag-of-words inference really is not that unmotivated, theoretically. I have shown that bag-of-words inference fits perfectly well into the logical scheme of things.
- And this theoretical account for the success of bag-of-words inference, especially when it comes to its robustness properties, is an important first step towards replicating the same robustness properties for other systems.

Outline

Computation via the Monte Carlo Method

- So, we have a unified theory that integrates deep and shallow inference. But an important question remains. Can we come up with a method, putting this into practice? How could we go about translating these theoretical informativity and robustness properties into an observable accuracy figure?
- And, as I've said before, I have only recently started to address this question in my work, and cannot yet make any concrete claims.
- But, nevertheless, I'd like to show you how I plan to approach he problem, and why this approach intuitively looks so interesting to me.



Model Theory: Satisfiability, Validity, Expectation

Definition

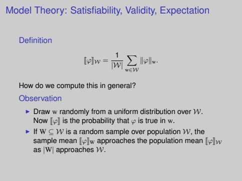
$$[\![\varphi]\!]_{\mathcal{W}} = \frac{1}{|\mathcal{W}|} \sum_{\mathrm{w} \in \mathcal{W}} \|\varphi\|_{\mathrm{w}}.$$

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 $\mathcal{W}$.

- The questin is, quite simply, how do we compute a degree of validity in practice?
- Obviously, an exact computation for this is far from trivial, since we have an exponential number of truth valuations to sum over, in this formula.
- Even, traditionally, if you want to check whether a propositional formula is a tautology, by trying out all truth assignments, you'll be in trouble computationally.
- This is because, when you have N propositions, you have to run your model-checker 2^N times, to ensure that all of the assignments of truth values make the formula true.
- This would be just too computationally complex, and that is the reason why theorem provers are usually symbolic implementations of the proof theory for a logic, not model checkers.



Model Theory: Satisfiability, Validity, Expectation

Definition

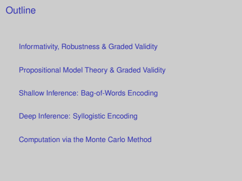
$$[\![\varphi]\!]_{\mathcal{W}} = \frac{1}{|\mathcal{W}|} \sum_{\mathrm{w} \in \mathcal{W}} \|\varphi\|_{\mathrm{w}}.$$

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 $\mathcal{W}$.

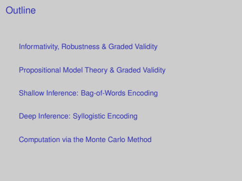
- But, fortunately, this arithmetic mean is a little better behaved than the strict notion of validity, when it comes to systematic model checking.
- Namely: By statistical sampling theory, we know that it has a consistent estimator.
- So if we only need to **approximate** this number, there is no need to check **all** of the valuations.
- We can simply check a **random sample**. And that is what enable the Monte Carlo method that I talked about about last year.
- That is: We can assign truth values to propositions at random, check the truth value, repeat the randomization, check the next truth value, repeat the randomization, etc. And we simply do this a couple of times and compute this mean over the sample rather than the whole population to get a reasonable estimate of the population mean.



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

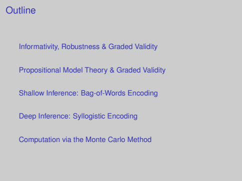
- I hope, that I’ve successfully made clear to you the idea of robust inference that I’m employing for my work, and some of its theoretical foundations within logic, combinatorics, and sampling theory.
- My theoretical framework is largely based on logic, however it advocates an important paradigm shift. I took the viewpoint that we have to introduce robustness properties to deal effectively with incomplete theories of background knowledge.
- This is done by generalizing from the classical dichotomous validity notion to a notion of graded validity, imposing a more useful structure on truth classes to deal with the case of what would traditionally be considered incomplete or inconsistent background theories.



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

- I have then shown that this theory is in fact a useful way to think and talk about current deep and shallow textual inference systems.
- On the shallow end of the spectrum: If you give me a logical representation for a bit of text, that comes directly out of a tokeniser, my notion of graded validity reduces to a bag-of-words overlap metric.
- If you give me a logical representation that comes out of deep analysis, where you give me predicate-argument structures and quantifier scopes, I can do everything that a theorem prover can do, and I can give you certain robustness properties on top of that, in order to deal with the situation where you're missing background knowledge.



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

- I've defined a notion of robustness for inference in NLP that is theoretically justifiable from the point of view of epistemology, logic, and linguistics.
- It enables practical computation of degrees of validity via a Monte Carlo method.
- So thank you, for your attention, and I'll be happy to answer any questions.