Entailment Relations on Distributions

J. van de Wetering

Department of Computer Science
University of Oxford

Workshop on Semantic Spaces at the Intersection of NLP, Physics and Cognitive Science, 2016

Lexical Entailment and Disambiguation

Entailment

"Dog bites cat"	$\Longrightarrow$	"Animal bites cat"
NOT "Animal bites cat"	$\Longrightarrow$	"Dog bites cat"

"Dog" *entails* "Animal"

Lexical Entailment and Disambiguation

Entailment

	"Dog bites cat"	$\Longrightarrow$	"Animal bites cat"
NOT	"Animal bites cat"	$\Longrightarrow$	"Dog bites cat"

"Dog" *entails* "Animal"

Disambiguation

bank	$\Longrightarrow$	river bank investment bank
cut	$\Longrightarrow$	cut of a movie a type of wound

Lexical Entailment and Disambiguation (2)

Properties:

- Reflexive
- Not symmetric
- Transitive*

Lexical Entailment and Disambiguation (2)

Properties:

- Reflexive
- Not symmetric
- Transitive*
- Related to information content

Lexical Entailment and Disambiguation (2)

Properties:

- Reflexive
- Not symmetric
- Transitive*
- Related to information content

$\implies$ Try to capture this in a partial order on distributions

Information Ordering

Definition

An information ordering is a partial order on distributions with suitable information like properties

Information Ordering

Definition

An information ordering is a partial order on distributions with suitable information like properties

Partial Order

A relation $\sqsubseteq$ such that for all x, y and z

- Reflexive: $x \sqsubseteq x$
- Transitive: $x \sqsubseteq y$ and $y \sqsubseteq z$ implies $x \sqsubseteq z$
- Anti-symmetric: $x \sqsubseteq y$ and $y \sqsubseteq x$ implies $x = y$

Information Ordering

Definition

An information ordering is a partial order on distributions with suitable information like properties

Partial Order

A relation $\sqsubseteq$ such that for all x, y and z

- Reflexive: $x \sqsubseteq x$
- Transitive: $x \sqsubseteq y$ and $y \sqsubseteq z$ implies $x \sqsubseteq z$
- Anti-symmetric: $x \sqsubseteq y$ and $y \sqsubseteq x$ implies $x = y$

Information Content

- Permutation invariance: $x \sqsubseteq y$ implies $\sigma(x) \sqsubseteq \sigma(y)$

Information Ordering

Definition

An information ordering is a partial order on distributions with suitable information like properties

Partial Order

A relation $\sqsubseteq$ such that for all x, y and z

- Reflexive: $x \sqsubseteq x$
- Transitive: $x \sqsubseteq y$ and $y \sqsubseteq z$ implies $x \sqsubseteq z$
- Anti-symmetric: $x \sqsubseteq y$ and $y \sqsubseteq x$ implies $x = y$

Information Content

- Permutation invariance: $x \sqsubseteq y$ implies $\sigma(x) \sqsubseteq \sigma(y)$
- Mixing: $x \sqsubseteq y$ implies $x \sqsubseteq tx + (1-t)y \sqsubseteq y$

Information Ordering

Definition

An information ordering is a partial order on distributions with suitable information like properties

Partial Order

A relation $\sqsubseteq$ such that for all x, y and z

- Reflexive: $x \sqsubseteq x$
- Transitive: $x \sqsubseteq y$ and $y \sqsubseteq z$ implies $x \sqsubseteq z$
- Anti-symmetric: $x \sqsubseteq y$ and $y \sqsubseteq x$ implies $x = y$

Information Content

- Permutation invariance: $x \sqsubseteq y$ implies $\sigma(x) \sqsubseteq \sigma(y)$
- Mixing: $x \sqsubseteq y$ implies $x \sqsubseteq tx + (1-t)y \sqsubseteq y$
- Minimal element: The uniform distribution $\perp_n = \frac{1}{n}(1, \ldots, 1)$

Information Ordering

Definition

An information ordering is a partial order on distributions with suitable information like properties

Partial Order

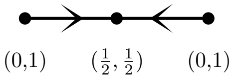
A relation $\sqsubseteq$ such that for all x, y and z

- Reflexive: $x \sqsubseteq x$
- Transitive: $x \sqsubseteq y$ and $y \sqsubseteq z$ implies $x \sqsubseteq z$
- Anti-symmetric: $x \sqsubseteq y$ and $y \sqsubseteq x$ implies $x = y$

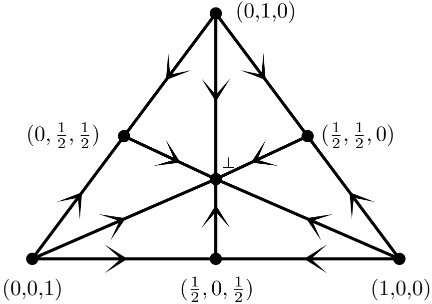
Information Content

- Permutation invariance: $x \sqsubseteq y$ implies $\sigma(x) \sqsubseteq \sigma(y)$
- Mixing: $x \sqsubseteq y$ implies $x \sqsubseteq tx + (1-t)y \sqsubseteq y$
- Minimal element: The uniform distribution $\bot_n = \frac{1}{n}(1, \ldots, 1)$
- Maximal elements: Pure states $\top_i = (0, \ldots, 0, 1, 0 \ldots, 0)$

Information flow in low dimensions



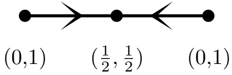
(a) Information flow for $n = 2$



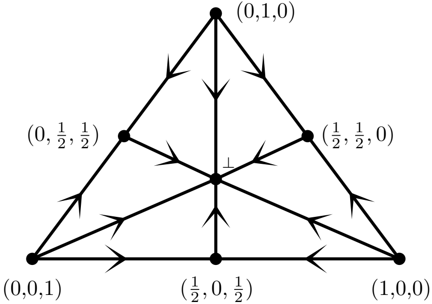
(b) Information flow for $n = 3$

- Information order is unique for $n = 2$

Information flow in low dimensions



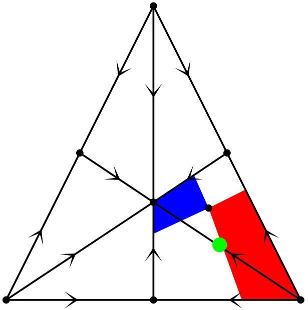
(a) Information flow for $n = 2$



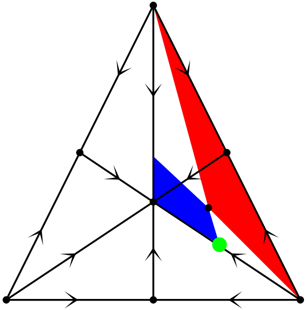
(b) Information flow for $n = 3$

- Information order is unique for $n = 2$
- Not unique for $n = 3$ and higher!

Renormalised Löwner orders

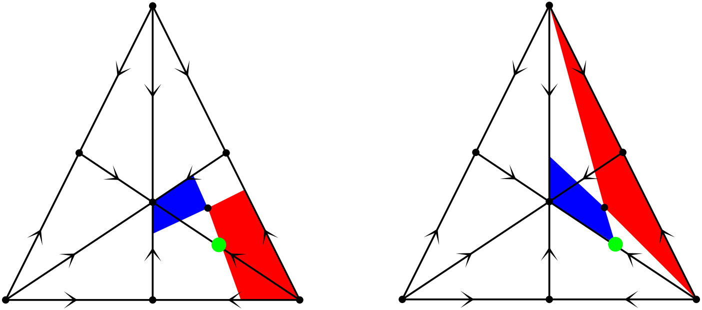


(a) $\sqsubseteq_L^+$: Renormalised to highest eigenvalue



(b) $\sqsubseteq_L^-$: Renormalised to lowest eigenvalue

Renormalised Löwner orders



(a) $\sqsubseteq_L^+$: Renormalised to highest eigenvalue

(b) $\sqsubseteq_L^-$: Renormalised to lowest eigenvalue

If $y = \frac{1}{30}(15, 10, 5)$ and $x = \frac{1}{10}(6, 2, 2)$
Then: $y \sqsubseteq_L^+ x$ and $x \sqsubseteq_L^- y$

Restricted Information Orders

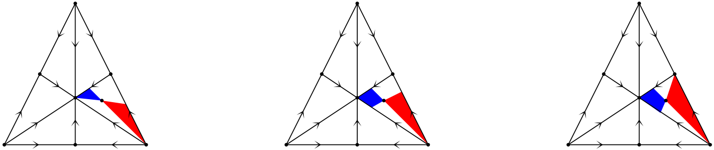
- An extra restriction: The *degeneracy* condition
- Makes sure comparisons are restricted to within sectors

Restricted Information Orders

- An extra restriction: The *degeneracy* condition
- Makes sure comparisons are restricted to within sectors
- $O(n^2)$ free parameters
- Set of restricted orders is a complete lattice

Restricted Information Orders

- An extra restriction: The *degeneracy* condition
- Makes sure comparisons are restricted to within sectors
- $O(n^2)$ free parameters
- Set of restricted orders is a complete lattice



(a) The minimal order (b) Bayesian order (c) The maximal order

Restricted Information Orders (2)

- No contradiction with $\sqsubseteq_L^+$ and $\sqsubseteq_I^-$
- Restricted information orders might prove too... restrictive

Restricted Information Orders (2)

- No contradiction with $\sqsubseteq_L^+$ and $\sqsubseteq_I^-$
- Restricted information orders might prove too... restrictive
- There is a natural way to define grading
- Other techniques for smoothing/grading

Summary

- There is an analog between entailment/disambiguation and information content
- There is a large selection of possible partial orders that could encode this
- Empirical research needed
- Compositionality?