Entailment in vector-space models

Chris Potts, Ling 236/Psych 236c: Representations of meaning, Spring 2013

May 23

1 Conceptualizing the problem

(1) Which row vectors entail which others?

	d_1	d_2	d_3
w_1	1	0	0
w_2	0	0	10
w_3	0	0	20
w_4	0	10	10
w_5	20	20	20

Possible criteria:

- Subset relationship on environments/features
- Score sizes
- Similarity of score vectors
- ...

2 Measures (based on Kotlerman et al. 2010)

Definition 1 (Feature functions)**.** Let u be a vector of dimension n. Then F_u is the partial function from $[1, n]$ such that $F_u(i)$ is defined iff $1 \leqslant i \leqslant n$ and $u_i > 0$. Where defined, $F_u(i) = u_i$.

Definition 2 (Feature function membership)**.** $i \in F_u$ iff i is defined for F_u

Definition 3 (Feature function intersection)**.** $F_u \cap F_v = \{i : i \in F_u \text{ and } i \in F_v\}$

Definition 4 (Feature function cardinality)**.** $|F_u| = \big|\{i : i \in F_u\}\big|$

Definition 5 (Weeds & Weir 2003)**.** $\mathit{WeedsPrec}(u, v) \stackrel{def}{=} \frac{\sum_{i \in F_u \cap F_v} F_u(i)}{\sum_{i \in F_u} F_u(i)}$

Definition 6 (Clarke 2009)**.** $\mathit{ClarkeDE}(u, v) \stackrel{def}{=} \frac{\sum_{i \in F_u \cap F_v} \min\left(F_u(i), F_v(i)\right)}{\sum_{i \in F_u} F_u(i)}$

Definition 7 (Kotlerman et al. 2010)**.**

$$\mathit{APinc}(u, v) \stackrel{def}{=} \frac{\sum_{i \in F_u} P(i) \cdot \mathit{rel}(F_r)}{|F_v|}$$

i. $\mathit{rank}(i, F_u) =$ the rank of $F_u(i)$ according to the value of $F_u(i)$

ii. $P(i) = \frac{\left|\{j \in F_v : \mathit{rank}(j, F_u) \leqslant \mathit{rank}(i, F_u)\}\right|}{\mathit{rank}(i, F_u)}$

iii. $\mathit{rel}(i) = \begin{cases} 1 - \frac{\mathit{rank}(i, F_v)}{|F_v|+1} & \text{if } i \in F_v \\ 0 & \text{if } i \notin F_v \end{cases}$

Definition 8 (Lin 1998). $LIN(u,v) \stackrel{def}{=} \frac{\sum_{i \in F_u \cap F_v} F_u(i)+F_v(i)}{\sum_{i \in F_u} F_u(i)+\sum_{i \in F_v} F_v(i)}$

Definition 9 (Kotlerman et al. 2010). If $E \in \{WeedsPrec, ClarkeDE, APinc\}$, then

$$balE(u,v) \stackrel{def}{=} \sqrt{LIN(u,v) \cdot E(u,v)}$$

2.1 Examples based on (1)

Maximal values are highlighted. Entailment testing is from row to column.

	w_1	w_2	w_3	w_4	w_5
w_1	1.0	0.0	0.0	0.0	1.0
w_2	0.0	1.0	1.0	1.0	1.0
w_3	0.0	1.0	1.0	1.0	1.0
w_4	0.0	0.5	0.5	1.0	1.0
w_5	0.3	0.3	0.3	0.7	1.0

(a) *WeedsPrec*

	w_1	w_2	w_3	w_4	w_5
w_1	1.0	0.0	0.0	0.0	0.6
w_2	0.0	1.0	1.0	0.8	0.7
w_3	0.0	1.0	1.0	0.9	0.7
w_4	0.0	0.6	0.6	1.0	0.9
w_5	0.3	0.4	0.4	0.7	1.0

(b) *balWeedsPrec*

Table 1: *WeedsPrec* and *balWeedsPrec*.

	w_1	w_2	w_3	w_4	w_5
w_1	1.0	0.0	0.0	0.0	1.0
w_2	0.0	1.0	1.0	1.0	1.0
w_3	0.0	0.5	1.0	0.5	1.0
w_4	0.0	0.5	0.5	1.0	1.0
w_5	0.0	0.2	0.3	0.3	1.0

(a) *ClarkeDE*

	w_1	w_2	w_3	w_4	w_5
w_1	1.0	0.0	0.0	0.0	0.6
w_2	0.0	1.0	1.0	0.8	0.7
w_3	0.0	0.7	1.0	0.6	0.7
w_4	0.0	0.6	0.6	1.0	0.9
w_5	0.1	0.3	0.4	0.5	1.0

(b) *balClarkeDE*

Table 2: *ClarkeDE* and *balClarkeDE*.

	w_1	w_2	w_3	w_4	w_5
w_1	0.5	0.0	0.0	0.0	0.2
w_2	0.0	0.5	0.5	0.2	0.1
w_3	0.0	0.5	0.5	0.2	0.1
w_4	0.0	0.2	0.2	0.5	0.2
w_5	0.5	0.2	0.2	0.3	0.5

(a) *APinc*

	w_1	w_2	w_3	w_4	w_5
w_1	0.7	0.0	0.0	0.0	0.3
w_2	0.0	0.7	0.7	0.3	0.2
w_3	0.0	0.7	0.7	0.4	0.2
w_4	0.0	0.4	0.4	0.7	0.4
w_5	0.4	0.3	0.3	0.5	0.7

(b) *balAPinc*

Table 3: *APinc* and *balAPinc*.

3 Baroni et al. (2012)

3.1 Entailment between nouns

	Relationship	Size
Positive class	$A\ N \models N$	1246 pairs
Negative class	$A\ N_2 \not\models N_1$	1246 pairs

(a) Training data.

	Relationship	Size
Positive class	$N_1 \models N_2$	1385 pairs, from WordNet hypernym chains
Negative class	$N_1 \not\models N_2$	1385 pairs, by inverting and shuffling the positive pairs

(b) Test data.

Table 4: All the data were manually checked after generation, and all the phrase types have at least 100 tokens in their data.

3.2 Entailment between quantified NPs

	Relationship	Size
Positive class	$Q_1\ N \models Q_2\ N$	7537 pairs; $Q_1 \models Q_2$ determined by hand
Negative class	$Q_1\ N \not\models Q_2\ N$	8455 pairs; $Q_1 \not\models Q_2$ determined by hand

(a) Data.

'Pair out'	Train on all but one quantifier pairs, assess on the remaining one
'Quantifier out'	Train on all the quantifiers, assess on the remaining one

(b) Experimental set-ups.

Table 5: All the data were manually checked after generation, and all the phrase types have at least 100 tokens in their data.

3.3 Unsupervised method

The authors use *balAPinc* as defined above and find that it beats their frequency- and similarity-based baselines on the nouns task but that it performs poorly on the quantifier task. (See page 30 for details on the performance and the thresholds used to define entailment categorically.)

3.4 Supervised method

- In the supervised approach, the authors train Support Vector Machines (SVMs) on concatenation of vector representations, reduced to 300 each dimensions with SVD/LSA.
- Their SVMs have polynomial kernels that captures feature interactions (p. 29).
- This method is successful for both the nouns task and the quantifiers task (Tables 3, 4).
- In the 'quantifier-out' set-up, performance ranges from 34% accuracy (*either*) to 98% (*each*).
- In addition, they tried working with just quantifier vectors (no N complements) and judged the model unsuccessful (p. 30).

References

Baroni, Marco, Raffaella Bernardi, Ngoc-Quynh Do & Chung-chieh Shan. 2012. Entailment above the word level in distributional semantics. In *Proceedings of the 13th conference of the European chapter of the association for computational linguistics*, 23–32. Avignon, France: ACL.

Clarke, Daoud. 2009. Context-theoretic semantics for natural language: An overview. In *Proceedings of the workshop on geometrical models of natural language semantics*, 112–119. Athens, Greece: ACL.

Kotlerman, Lili, Ido Dagan, Idan Szpektor & Maayan Zhitomirsky-geffet. 2010. Directional distributional similarity for lexical inference. *Natural Language Engineering* 16(4). 359–389. doi: 10.1017/S1351324910000124.

Lin, Dekang. 1998. Automatic retrieval and clustering of similar words. In *Proceedings of COLING-ACL*, 768–774. Montreal: ACl.

Weeds, Julie & David Weir. 2003. A general framework for distributional similarity. In Michael Collins & Mark Steedman (eds.), *Proceedings of the 2003 conference on empirical methods in natural language processing*, 81–88.