Theoretical study and empirical investigation of sentence analogies

Stergos Afantenos[1] Suryani Lim[2] Henri Prade[1] Gilles Richard[1]

[1]IRIT, Toulouse, France

[2]Federation University, Churchill, Australia

July 2022

Analogies: from words to sentences

1. Analogical reasoning: essential component of human intelligence (Aristotle, etc.)
2. Gentner *et al*, Hofstadter *et al*, etc.
3. Analogical proportions: building block for analogical reasoning
4. Parallel between 2 situations (a, b) and (c, d) $a : b :: c : d$

   Vienna (a) is to Austria (b) as Paris (c) is to France (d)

Analogies: from words to sentences

1. Analogical reasoning: essential component of human intelligence (Aristotle, etc.)
2. Gentner *et al*, Hofstadter *et al*, etc.
3. Analogical proportions: building block for analogical reasoning
4. Parallel between 2 situations (a, b) and (c, d) $a : b :: c : d$

   Vienna (a) is to Austria (b) as Paris (c) is to France (d)
5. But what about:
   John sneezed loudly (a). Mary was startled (b).
   Bob took an analgesic (c). His headache stopped (d).

How to make it clear ?

First remember some postulates...

The main ones:

1. $a : b :: a : b$ (*reflexivity*)
2. $a : b :: c : d \rightarrow c : d :: a : b$ (*symmetry*)
3. $a : b :: c : d \rightarrow a : c :: b : d$ (*central permutation*)

First remember some postulates...

The main ones:

1. $a : b :: a : b$ (*reflexivity*)
2. $a : b :: c : d \rightarrow c : d :: a : b$ (*symmetry*)
3. $a : b :: c : d \rightarrow a : c :: b : d$ (*central permutation*)

With immediate consequences:

- $a : a :: b : b$ (*identity*);
- $a : b :: c : d \rightarrow b : a :: d : c$ (*internal reversal*);
- $a : b :: c : d \rightarrow d : b :: c : a$ (*extreme permutation*);
- $a : b :: c : d \rightarrow d : c :: b : a$ (*complete reversal*).

24 permutations $\Longrightarrow$ 3 classes of 8 equivalent proportions

But not working for sentences...

What is the issue?

OK: first sentence in a pair entails the second one

John sneezed loudly (a). Mary was startled (b).
Bob took an analgesic (c). His headache stopped (d).

NO WAY: no clear relation between first sentence and second one

John sneezed loudly (a). Bob took an analgesic (c).
Mary was startled (b). His headache stopped (d).

Central permutation not relevant ?

A proposal

Switch $a : b :: c : d \rightarrow a : c :: b : d$ (*central permutation*)
for $a : b :: c : d \rightarrow b : a :: d : c$ (*internal reversal*)
Moving to "weak analogical proportion":

1. $a : b :: a : b$ (*reflexivity*)
2. $a : b :: c : d \rightarrow c : d :: a : b$ (*symmetry*)
3. $a : b :: c : d \rightarrow b : a :: d : c$ (*internal reversal*)

Back to our sentence example:

> **Mary was startled (b) because John sneezed loudly (a).**
> **His headache stopped (d) because Bob took an analgesic (c).**

Entailment replaced by passive form: highly acceptable!

Question: Can we experimentally show under which conditions internal reversal holds?

Experiments

- Four series of experiments in order to investigate internal reversal empirically
- With following three datasets:
  - **PDTB:** contains annotated pairs of sentences with either implicit or explicit discourse relations.
  - **SNLI:** contains pairs of sentences annotated with *entailment, contradiction* or *neutral*.
  - **MSRPC:** contains pairs of sentences annotated as paraphrases or not.
- For each dataset, two random pairs chosen; if they were annotated with the same relation the quadruplet was marked as a positive instance, negative otherwise.
- Positive and negative instances well-balanced (2M instances for each class).

Experimental settings

Base setting

- **Training set:** $(\mathscr{X}_{train}, \mathscr{Y}_{train}) = (\{\mathbf{x}_i\}_{i=1}^{n}, \{y_i\}_{i=1}^{n})$
- **Test set:** $(\mathscr{X}_{test}, \mathscr{Y}_{test}) = (\{\mathbf{x}_i\}_{i=1}^{m}, \{y_i\}_{i=1}^{m})$
- m is a tenth of n
- $\mathbf{x}_j$ represents a quadruplet of sentences $a : b :: c : d$ with $y_j = 1$ if (a, b) and (c, d) share the same relation, otherwise $y_j = 0$
- The training and test sets well balanced : $|\{y_k : y_k = 1\}| = |\{y_k : y_k = 0\}|$

From training set we learn a model $\mathscr{H}_b$ which we can test on the test set.

Experimental settings

Internal reversal on the test set (exp. set. 1)

- **Training set:** same as base setting
- **Test set:** internal reversal on test set used in base model

Question: under which conditions do we get similar results as base model?

Experimental settings

Test set from train distribution with internal reversal (exp. set. 2)

- **Training set:** same as the base model
- **Test set:** for every positive instance $(\mathbf{x}_{a:b::c:d}, 1) \in (\mathscr{X}_{train}, \mathscr{Y}_{train})$ we add *internal reversal* pair $(\mathbf{x}_{b:a::d:c}, 1)$ to a new test set $(\mathscr{X}_{test}, \mathscr{Y}_{test})$ whose size thus is $n/2$.

Question: how well does a trained model can detect analogies after performing internal reversal on the same set of pairs of sentences.

Experimental settings

Augmenting training and test sets (exp. set. 3)

- **Training set:** $(\mathscr{X}^a_{train}, \mathscr{Y}^a_{train}) = (\{\mathbf{x}_i\}_{i=1}^{n+n/2}, \{y_i\}_{i=1}^{n+n/2}\})$
- **Test set:** $(\mathscr{X}^a_{test}, \mathscr{Y}^a_{test}) = (\{\mathbf{x}_i\}_{i=1}^{m+m/2}, \{y_i\}_{i=1}^{m+m/2})$
- For both sets we start from base train and test sets and for all $(\mathbf{x}_{a:b::c:d}, 1)$ we further add internal reversal $(\mathbf{x}_{b:a::d:c}, 1)$.

Experimental settings

Augmenting test set (exp. set. 4)

- **Training set:** same as base model
- **Test set:** $(\mathscr{X}_{test}^{a_t}, \mathscr{Y}_{test}^{a_t}) = (\{\mathbf{x}_i\}_{i=1}^m, \{y_i\}_{i=1}^m)$ by keeping only positive instances of the base training set, which is then augmented with instances $(\mathbf{x}_{b:a::d:c}, 1)$ for every instance $(\mathbf{x}_{a:b::c:d}, 1)$, yielding thus same size.

Random Forests Model

Parameters used for Random Forests:

- 100 trees
- no maximum depth
- minimum split of two

Bi-LSTM Model

- Input: a set of quadruplets of sentences $\{a : b :: c : d\}_{i=1}^{n}$ associated with $y_{i=1}^{n} \in \{0, 1\}$
- Each sentence composed of tokens $s = \{w_1^s, \ldots, w_{|s|}^s\}$ with $s \in \{a, b, c, d\}$
- Maximum length of sentences 35 words (padding for shorter sentences)
- we obtain representation $s = \overrightarrow{h^s} \mathbin{+\!\!+} \overleftarrow{h^s}$ for each sentence $s \in \{a, b, c, d\}$ with $\mathbin{+\!\!+}$ representing concatenation.

Bi-LSTM Model (cont.)

- We obtain

$$\mathbf{h}_f = f(\mathbf{W}^T \mathbf{h}_{LSTM} + \mathbf{b})$$

with

$$\mathbf{h}_{LSTM} = \overrightarrow{h^a} + \overleftarrow{h^a} + \overrightarrow{h^b} + \overleftarrow{h^b} + \overrightarrow{h^c} + \overleftarrow{h^c} + \overrightarrow{h^d} + \overleftarrow{h^d}$$

- Finally we give this as input to an MLP

$$\mathbf{h}_f = f(\mathbf{W}^T \mathbf{h}_{LSTM} + \mathbf{b})$$

- Final layer contains a sigmoid function - Binary Cross Entropy as Loss function
- GloVe vectors (dimension 300) as input embeddings for LSTM.

Results

LSTM (left) Random Forest (right)

LSTM

		Precision	Recall	F1	Accuracy
	PDTB				
base setting	class 1	54.274	47.476	50.648	53.739
	class 0	53.322	60.001	56.465	
Exp. Set. 1	class 1	48.91	39.76	43.863	49.114
	class 0	49.254	58.468	53.467	
Exp. Set. 2	class 1	100.0	39.76	56.898	39.76
Exp. Set. 3	class 1	70.16	79.346	74.471	63.733
	class 0	44.038	32.507	37.404	
Exp. Set. 4	class 1	100.0	46.585	63.56	46.585
	SNLI				
base setting	class 1	67.862	67.811	67.837	67.859
	class 0	67.856	67.907	67.882	
Exp. Set. 1	class 1	50.111	49.57	49.839	50.11
	class 0	50.11	50.651	50.379	
Exp. Set 2	class 1	100.0	49.57	66.283	49.57
Exp. Set 3	class 1	84.489	83.982	84.235	79.047
	class 0	68.365	69.185	68.772	
Exp. Set. 4	class 1	100.0	59.086	74.282	59.086
	MRPC				
base setting	class 1	53.45	61.487	57.188	53.969
	class 0	54.671	46.45	50.227	
Exp. Set. 1	class 1	80.454	87.638	83.892	83.173
	class 0	86.426	78.708	82.387	
Exp. Set. 2	class 1	100.0	87.638	**93.412**	87.638
Exp. Set. 3	class 1	69.033	72.395	70.674	59.946
	class 0	38.832	35.05	36.844	
Exp. Set. 4	class 1	100.0	62.752	77.114	62.75

Random Forest

		Precision	Recall	F1	Accuracy
	PDTB				
base setting	class 1	54.604	33.778	41.737	53.314
	class 0	52.744	72.468	61.053	
Exp. Set. 1	class 1	51.254	31.096	38.708	50.826
	class 0	50.639	70.504	58.943	
Exp. Set. 2	class 1	100.00	31.096	47.440	31.096
Exp. Set. 3	class 1	66.263	99.953	79.694	66.267
	class 0	69.847	0.213	0.424	
Exp. Set. 4	class 1	100.00	32.117	48.619	32.117
	SNLI				
base setting	class 1	50.725	47.006	48.794	50.729
	class 0	50.732	54.443	52.522	
Exp. Set. 1	class 1	50.302	46.189	48.158	50.285
	class 0	50.270	54.379	52.244	
Exp. Set 2	class 1	100.00	46.189	63.191	46.189
Exp. Set 3	class 1	70.368	86.903	77.766	66.898
	class 0	50.797	26.979	35.241	
Exp. Set. 4	class 1	100.00	46.313	63.307	46.313
	MRPC				
base setting	class 1	54.327	69.353	60.927	54.739
	class 0	55.502	39.599	46.221	
Exp. Set. 1	class 1	58.916	77.847	67.071	61.374
	class 0	66.313	44.547	53.293	
Exp. Set. 2	class 1	100.00	77.847	**87.544**	77.847
Exp. Set. 3	class 1	67.523	99.649	80.499	67.437
	class 0	48.952	0.698	1.377	
Exp. Set. 4	class 1	100.0	69.139	81.754	69.139

Conclusion

- **Experimental setting 1**: When the relation is symmetrical (e.g. paraphrases), better results with internal reversal compared to base setting
- **Experimental setting 2**: Focuses on positive instances on test set and the results are similar (better for symmetrical relations)
- **Experimental setting 3**: Results universally better
- **Experimental setting 4**: Focuses on positive instances - results better for symmetrical relations only.