Learning to compose spatial relations with grounded neural language models

Mehdi Ghanimifard
mehdi.ghanimifard@gu.se
University of Gothenburg, Sweden

Simon Dobnik
simon.dobnik@gu.se
University of Gothenburg, Sweden

Aims

- A probabilistic model for learning grounded spatial templates from examples of language use.
- Evaluate conditional neural language model for grounded semantic composition.
- Synthetic dataset of language use of simple and composite spatial relations based on Logan and Sadler (1996).

Spatial Templates

- Spatial templates are representations of regions of acceptability with aligned frame of reference associated with a spatial relation, centered on reference object.

Grounded Neural Language Model

- Simple Language Model: repeatedly predict the next word.

$$Pr(w_{1:T}) = \prod_{t=1}^{T} Pr(w_t|w_{1:t-1})$$

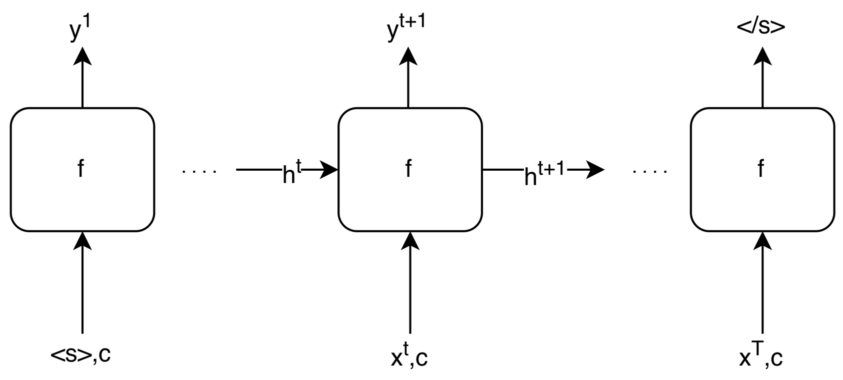
- Grounded Neural Language Model: conditioned language model under sensors state:

$$P(w_{1:t}|c) = \prod_{t=1}^{T} P(w_t|w_{1:t-1}, c)$$

- Recurrent Neural Language Model estimates parameters of a recurring function for next word probabilities in each step:

$$P(next\ word|w_{1:t-1}) = y_t$$
$$\hat{y}_t = softmax(f(w_{i-1}, f(w_{i-2}, f(\ldots, f(w_1)\ldots))))$$

- Add sensory data (location: c) in each time step to the language model.



As an optimization problem, with gradient based learning, parameters will be learned toward minimizing the categorical cross-entropy between predicted probability and delta distribution of observed samples. Similar to Graves (2013).

Lemma

- Degree of applicability scores as probabilities or degrees of belief; (Ramsey 1926) and (Coventry et al 2004)
- With the same argument:

$$Score(w_{1:T}, c) \propto Pr(w_{1:T}, c)$$
$$Pr(w_{1:T}, c) = Pr(w_{1:T}|c) \times Pr(c)$$

- By assuming that all locations on map are equally accessible, $Pr(c)$ is constant, then:

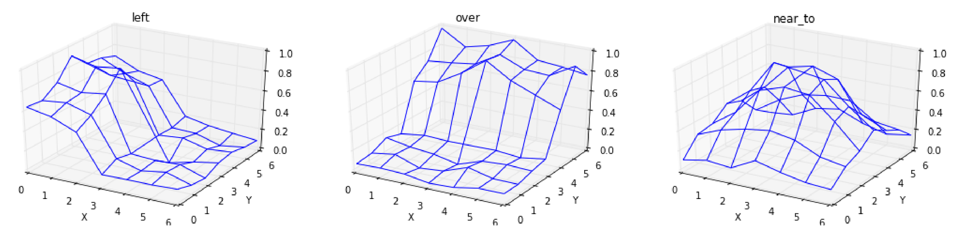
$$Score(w_{1:T}, c) \propto Pr(w_{1:T}|c)$$

This formula can be used for evaluation of the learned representation from language model comparing to human judgments.

Setup

Following the experimental setup of Logan and Sadler (1996)

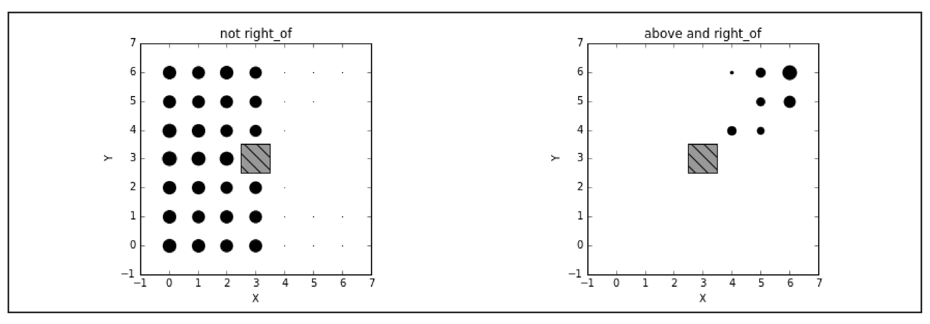
- *above, below, over, under, left of, right of, next to, away from, near to, far from*
- 7×7 space for their collected acceptability judgments.



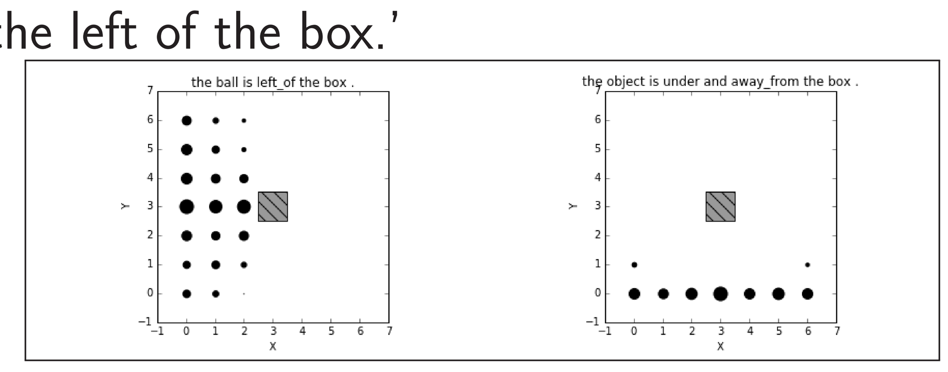
- We use a one layer vanila long short-term memory (LSTM) with an embedding layer and dropout.

Generating Synthetic Compositions

1. Negative compositions (e.g. *not right of*)
   Interesective compositions (e.g. *above and right of*)



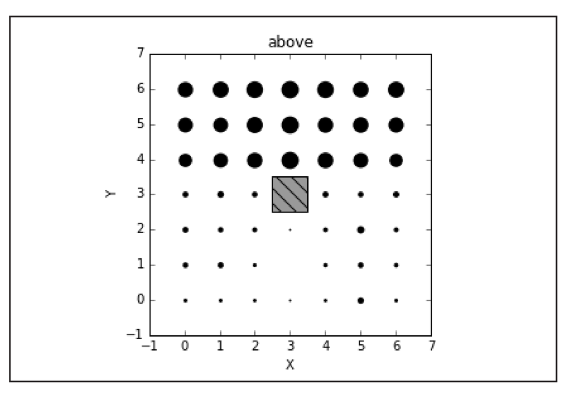
2. We added words such e.g. 'The object is to the left of the box.'

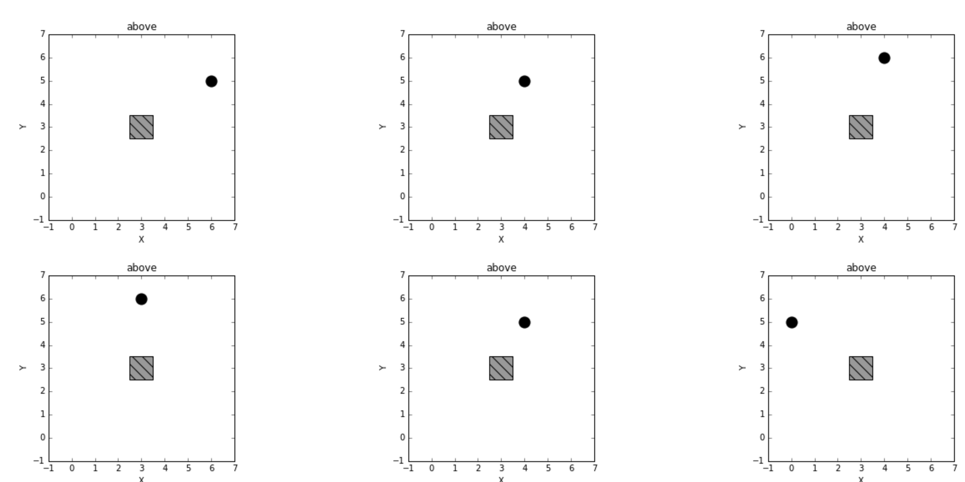


Training dataset

Language use:

- The training samples pairs of location and description (phrase). The frequency of each sample is based on the acceptability score. First we scale down all these *scores* between 1 and 9 to 0 and 1, then:
  $freq(phrase, c) = 100 \times score(phrase, c)$
- For example, 'above':

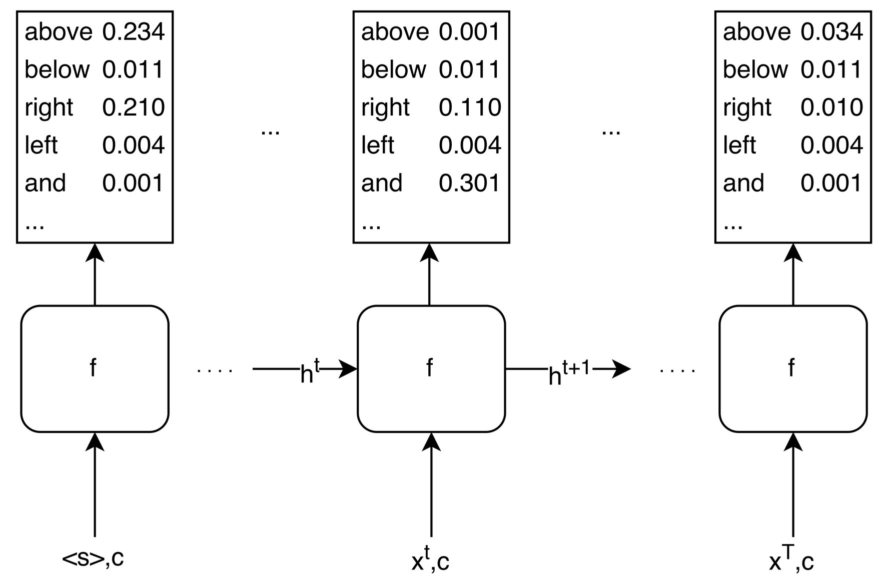




...

Example test outputs

- The model produces probabilities per time-step.

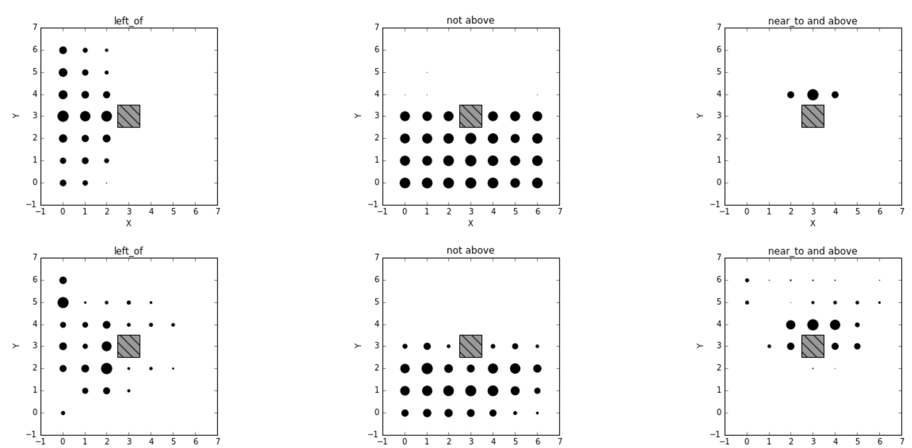


- Any input phrase, paired with a location c, provides probability of the phrase conditioned with c.
  Does the probability correlates with judgment scores?

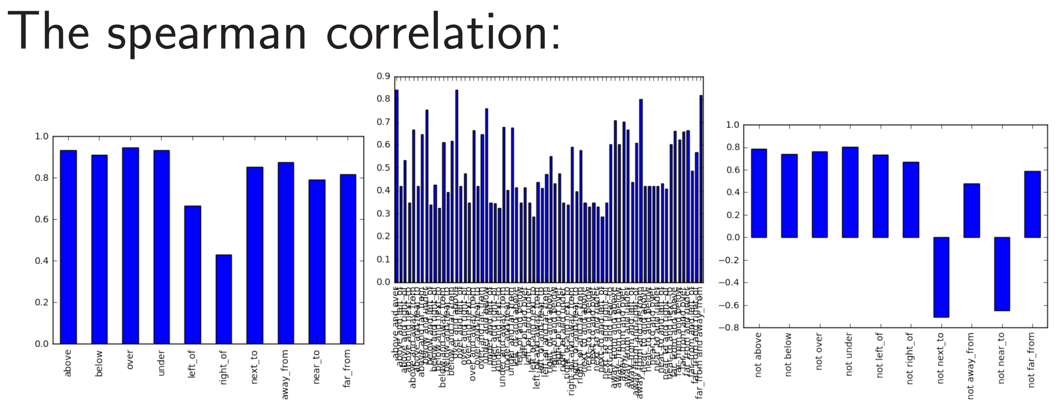
Evaluation and results

1. Before training, we randomly hold %10 of the corpus for test.

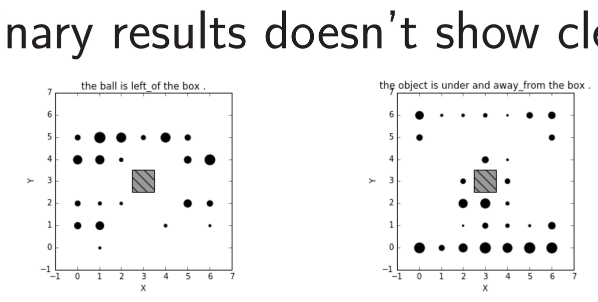
- Comparison: original (top) and the learned representations (bottom):



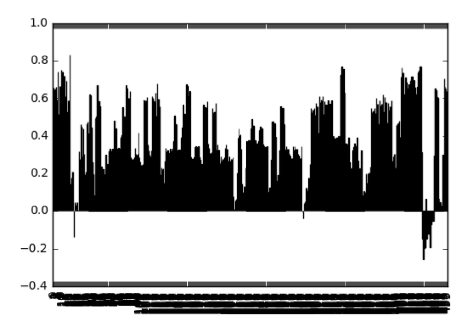
- The spearman correlation:



2. With the same evaluation setup, we examined the full sentence compositions. In this case, the number of parameters, are drastically higher and our preliminary results doesn't show clear success.



- The Spearman correlations for all sentences:



Conclusions and future work

- Neural language models can be used for modeling grounded meaning.
- Growing the non-grounded vocabulary makes it harder to converge to meaningful representation.
- Future work: expand our dataset with natural corpus, with more complicated constituent structure
- Explore transfer learning on word distributions for words not directly grounded.