Comparatives, Quantifiers, Proportions: A Multi-Task Model for the Learning of Quantities from Vision

Sandro Pezzelle\textsuperscript{1}, Ionut-Teodor Sorodoc\textsuperscript{2}, Raffaella Bernardi\textsuperscript{3}

\textsuperscript{1,2,3}CIMeC - Center for Mind/Brain Sciences, \textsuperscript{3}DISI, University of Trento

sandro.pezzelle@unitn.it | sandropezzelle.github.io | quantit-clic.github.io

\textbf{Motivation}

Quantification: Operation of expressing quantitative information

- ‘There are more cars than parking lots’: comparatives
- ‘Most of the supporters wear blue t-shirts’: quantifiers
- ‘20\% of the trees have been planted last year’: proportions
- ‘Seven students passed the exam’: numbers

Comparatives, Quantifiers, Proportions express a comparison or relation between sets; Numbers denote cardinality of one set

Different age of acquisition [1,2,3], no need of counting for using comparatives and quantifiers in grounded contexts [4]

\textbf{Hypothesis}

Cs, Qs, Ps express increasingly-complex steps of same ratio-based mechanism; Ns require different, possibly interfering operation [5]

\textbf{Task & Dataset}

Research Question
Can 3 ratio-based quantification tasks be modeled by a single, Multi-Task Learning (MTL) neural network from Vision?

\textbf{Multi-Task Learning Model}

In-Depth Evaluation

MTL outperforms one-task models: sharing weights helps!

MTL model approximates human data and makes ‘plausible’ errors

Results

Predictions

\textbf{In-Depth Evaluation}

\textbf{Numbers in the loop}
Introducing number of targets in the pipeline hurts performance!

\textbf{Reversing the architecture}
Proportions (.08 acc) > Quantifiers (.32 r) > Comparatives (.65 acc)

\textbf{Discussion & References}

Sharing a common core boosts performance in all tasks, proving their (a) interdependency and (b) increasing complexity

Are representations learned from one modality abstract enough to be transferable to different modalities, e.g. language, sounds?

\textbf{References}