Pietroski et al. (2009): two interpretation strategies for “most”

Cardinality-based strategy
most($A$, $B$) $\iff$ \[ |S_{A\cap B}| > \frac{1}{2} \cdot |A| \]
$\iff$ \[ |S_{A\cap B}| > |S_{A\backslash B}| \]
1. Estimate the number of entities satisfying both predicates ("red squares") and the number satisfying one predicate but not the other ("non-red squares").
2. Compare these number estimates and check whether the former is greater than the latter.

Pairing-based strategy
$A \leftrightarrow B \iff \forall x : A(x) \leftrightarrow B(x) \iff |S_A| = |S_B|$
most($A$, $B$) $\iff \exists S \subseteq S_{A\cup B}: S \leftrightarrow S_{A\cap B}$
1. Successively match entities satisfying both predicates ("red squares") uniquely with entities satisfying one predicate but not the other ("non-red squares").
2. The remaining entities are all of one type, so pick one and check whether it is of the first type ("red square").

$\mathbf{x}$: entity, $A$ and $B$: predicates (e.g., "square" and "red"), $A(x)$ true iff $x$ satisfies $A$, and $S_A = \{ x : A(x) \}$: set of entities satisfying $A$.

Examples

- Exactly two squares are yellow.
- Exactly no square is red.
- More than half the red shapes are squares.
- More than a third of the shapes are cyan.
- Less than half the shapes are green.
- Exactly all magenta shapes are squares.
- At most five shapes are magenta.
- At least one triangle is gray.

Three types of spatial arrangements

- Paired
- Random
- Partitioned

"More than half the shapes are red squares?"

Increasingly balanced attribute ratios

GitHub projects & PDF versions

ShapeWorld: https://github.com/AlexKuhnle/ShapeWorld
FiLM for ShapeWorld: https://github.com/AlexKuhnle/film
Paper & poster PDF, plus related papers: https://www.cl.cam.ac.uk/~aok25/

Experimental setup: task, model, data, etc

Task: image caption agreement
Model: FiLM (Perez et al. 2018)
Variants: pre indicates pretrained CNN module, -coll indicates shape collisions allowed
Training data: 100k images with 5 captions per image
Training: 100k iterations with batch size 64 (~ 13 epochs)
Validation data: 20k instances
Test data: 48 configurations with 1024 instances each

Numbers: "zero" to "five"
Quantifiers: "no", "a/ three quarter(s)", "a/two third(s)", "all"
Modifiers: "less than", "at most", "exactly", "at least", "more than", "not"
Training datasets: Q-full contains all quantifiers, Q-half contains only "more than half" and "less than half"
Test datasets: One contrasting attribute, close-to-balanced contrast attribute ratios, area- vs size-controlled, random/paired/partitioned positioning

Experimental results

Training performance

Evaluation performance

Weber fraction: performance for increasingly balanced ratios

Q-full model performance for increasingly balanced ratios (x-axis indicates ratio via n:n+1)

Performance as a function of the actual ratio fraction (n+1)/n, with Weber fraction (75%) highlighted

Weber fraction: performance for increasingly balanced ratios