Generating Expressions that Refer to Visible Objects

Margaret Mitchell, Kees van Deemter and Ehud Reiter (2013)

José Miguel Cano Santín

Visible Objects Algorithm: goals

- Generates descriptive, human-like referring expressions (REG) for visible objects.
- Identify a referent in an image to the listener.

TUNA corpus

Primary contributions

1. **Overspecification**/redundancy, **underspecification**.
    - Speakers not always select only the properties that have a contrastive value.
        - “The grey desk”
        - Overspecification: “The large grey desk”.
        - Underspecification: “The large desk”.
    - What is **salient**:
        - Visually: what does the visual system first respond to, what guides attention?
        - Linguistically: what do people tend to mention in visual scenes?
        - Cognitively: what is atypical for this object? (Not discussed)
    - Visually and linguistically salient features tend to be the same ones.
        - People tend to use color and size first when identifying objects.
        - These features are **favored** in the REG of this algorithm.

Primary contributions

2. Separation between:
   - **Absolute properties**
     - **Color**, shape and material.
     - May be detected directly using Computer Vision (CV) techniques.
       - (color: yellow)

   - **Relative properties**
     - **Size**, location and orientation.
     - Require reasoning over visual features to determine an appropriate form.
     - e.g. How can the system determine that an object from an image is big/small, tall/short, etc.?
       - Incorporation of the top-performing size algorithm introduced in Mitchell et al. (2011), which takes as **input** the **height** and **widths** of objects in the image and **outputs** a **size value** or **NONE**, indicating that size should not be used to describe the object.

Primary contributions

3. Stochastic nature of RE.
   - Speakers produce different references to the same object.

4. Evaluation method.
   - Non-deterministic REG
     - The algorithm will not return the same output every time.

The Visible Objects Algorithm

- Requirements.
  1. Prior likelihood ($\boldsymbol{\alpha_{att}}$).
     - Likelihood that an attribute (att) generates a corresponding visual property.
  2. Ordered list of absolute attributes beyond color (**AP**).
     - Empty for the evaluation corpora of this paper.
  3. Ordered list of relative attributes beyond size (**RP**).
     - Location and orientation.
  4. Ordered list of all attributes (**P**).
     - To stablish a preference order among all attributes.
  5. Ordered list specifying the order in which to scan through other scene objects.
     - The current implementation uses the order in which the objects are listed in the corpora it is run on.

The Visible Objects Algorithm

- The **Stochastic** Process.
  1. **Encourage the attributes** that we know people are likely to use.
     - Using its α_{att} as an estimate of whether to include it.

  2. **Penalize longer** descriptions.
     - With a penalty proportional to the length of the property set under construction.
     - Length-based penalty: γ.

$$f(A \cup \{x\}) = \gamma \alpha_{att}$$

$$\gamma = \begin{cases} \frac{1}{\lambda |A|} & \text{if } |A| > 0 \\ 1 & \text{otherwise} \end{cases}$$

The Visible Objects Algorithm

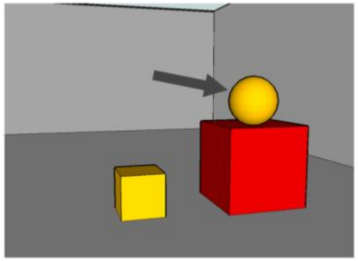
- **Scanning** Through Objects.
  - The algorithm compares each object in the scene that is the same type as the target.
  - If the values for an attribute are different, then the corresponding property is added to the property set based on the length penalty alone.
  - In development, it was found that incrementally scanning through objects resulted in better performance.

Evaluation

- Comparison with other algorithms.
  1. **Incremental Algorithm** (Dale and Reiter, 1995, as implemented in NLTK by Bird et al., 2009).
     - Iterates through attributes in a predefined preference order: color -> size -> type.
     - For each attribute, it checks whether specifying a value would rule out at least one item in the contrast set.
     - If it will, the (attribute: value) is added to the description.
  2. **Graph-Based Algorithm** (Krahmer et. al, 2003, as implemented in Viethen et al., 2008).
     - The objects in the discourse are represented within a labeled directed graph.
     - Each object is represented as a vertex.
       - Properties for an object represented as self-edges on the that vertex.
       - Spatial relations between objects represented as edges between vertices.
  - Both produce only one property set (output).

Evaluation corpora

1. **GRE3D3** corpus (Viethen and Dale, 2008).
   - Type, color, size and location.



2. **TUNA** corpus (van Deemter et al., 2006).
   - Type, color, size and orientation.

Evaluation metrics

1. **Aligned Dice**
   - Provides a value for the similarity between a generated description (*S*) and a human produced description (*H*).
   - For the corpus of *H* and the corpus of *S*, find the best alignment *x* out of all possible alignments *X* and apply the Dice function.

$$\arg\max_{x \in X} \sum_{(S,H) \in x} \mathrm{Dice}(D_S, D_H) \qquad \frac{|D_S \cap D_H|}{|D|}$$

2. **Majority.**
   - The proposed algorithm has more than one chance to match the human descriptions
   - Compares how often the most frequent generated set (*S*) compares with the most frequent observed set (*H*).
   - The majority score is the percentage of folds where these two sets match.

Evaluation methodology

1. The proposed algorithm is run 1,000 times.
2. The generated property sets are ordered by frequency.
3. The most frequent generated set is compared against the most frequent human produced.
4. The majority score is the percentage of folds where these two sets match.

- For IA and FB, the most frequent generated set is the only generated set.

Evaluation results

Algorithm	Aligned Dice		Majority	
	Set 1	Set 2	Set 1	Set 2
Proposed Alg.	**88.23**	**90.06**	62.50	**50.00**
IA	87.71	85.13	62.50	25.00
GB	87.71	88.73	62.50	**50.00**

Table 2: GRE3D3: Results (in %).

Algorithm	Aligned Dice		Majority	
	+LOC	-LOC	+LOC	-LOC
Proposed Alg.	**88.75**	**86.07**	**40.00**	40.00
IA	81.79	81.55	0.00	**100.00**
GB	75.36	66.04	20.00	20.00

Table 3: TUNA: Results (in %).