Improving Grounded Language Acquisition Efficiency using Interactive Labeling

Nisha Pillai, Karan K. Budhraja, Cynthia Matuszek
{ npillai1 | karanb1 | cmat } @ umbc.edu

Robot Interaction Task

2

- Goal: Interact using novel language about novel things
  - How: Interactively learn a joint model of percepts and language
- Robot's role
  - Learn to associate words with objects
  - Request for human assistance when required
  - Speech or text
- Human's role
  - Describe (annotate) or confirm annotations of objects
- Robot trains classifiers for attributes associated with words[1]
  - Novel percepts *and* novel language

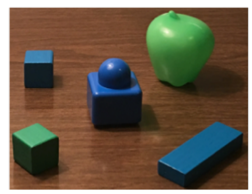
[1] Matuszek, FitzGerald, Zettlemoyer, Bo, Fox. ICML 2013.

Motivation

3

- Language is a comfortable, natural interaction mode
  - But you know that ☺
- Language is broad!
  - Too many domains and possibilities
  - Need (some) on-the-fly learning
  - Nobody likes annotation tasks – especially users
- This work: visual percepts ←→ attribute words
- Can active learning improve language acquisition efficiency and user acceptance?

"The apple"
"The Lego"
"The green round thing"



Learning Groundings

4

- Users describe objects the robot should understand
- When new language tokens are encountered
  - Visual classifiers are created and trained on the perceptual context
  - Tokens are associated with those classifiers
  - Create `NEW-CLASSIFIER-CALLED-'green'`
  - But also `NEW-CLASSIFIER-CALLED-'short'`, etc
- As more objects are seen, 'best' classifier emerges

Language Annotation

"This is a short green cube."

Perceived world state

Learning

Newly created semantics

Word "short"
←→
NEW-CLASSIFIER-CALLED-'short'
CALLED-'green'
CALLED-'cube'

Goals

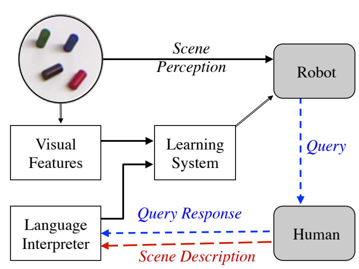
5

- Incorporate active learning in grounded language acquisition
- Improve learning *efficiency:*
  - Reduce the amount of annotation in labeling
- Improve learning *user-friendliness:*
  - Compare response to naïve annotation *vs.* interactive labeling
- Design an interactive, user-friendly model
- Compare two approaches:
  - Manual annotation: naïvely label entire corpus before training
  - Interactive labeling: provide labels or verification on request

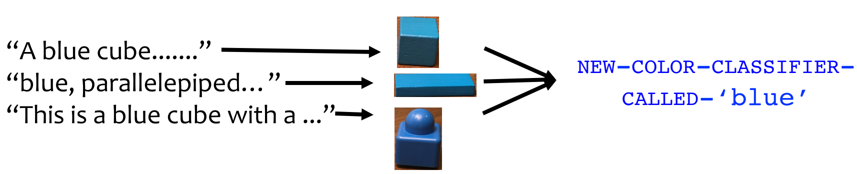
Interactive Labeling

6

- Query: Request description or verification
- Response: Verify or label (describe object)
  - Verification based on classifier confidence
- Interpreter: Extract keywords
- Visual Features: Segmented RGB values from point cloud



- Learning System: Train on all objects labeled with keyword



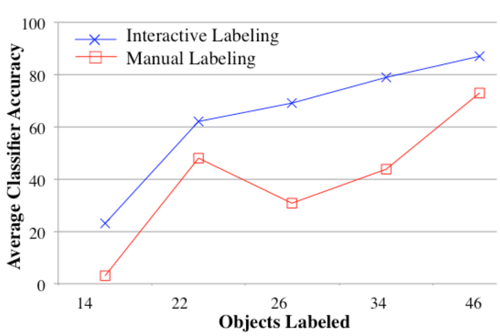
Experiments

7

- Learning Performance:
  - 240 distinct object views, 50 objects, 5 human annotators
  - Do classifiers improve faster in active learning?
    - E.g., effective learning from fewer annotations
- Comparative User Experience study:
  - 10 distinct object views, 10 objects, 10 participants
  - General (subjective) questions
  - Likert-scale technology acceptance questions

Classification Performance

8



- Performance of trained classifiers on test set
- Interactive labeling: fewer labels → same performance
- Main sources of error:
  - Perception!
  - Variation in shades of colors
  - Overfitting due to small set of training objects
  - Lighting conditions

Model Quality

9

- Performance of trained model
  - Ability to correctly classify held-out test set
- Goal: only classifiers associated with attribute keywords have strong confidence

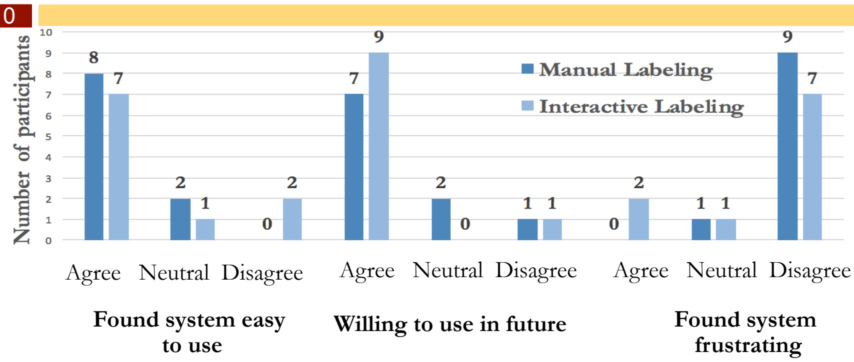
Classifiers

Ground truth	"arc"	"banana"	"blue"	"bottom"	"cylinder"	"green"	"half"
green	0.025	0.002	0.252	0.024	0.182	0.970	0.019
red	0.024	0.005	0.007	0.003	0.124	0.000	0.116
yellow	0.008	0.048	0.024	0.010	0.086	0.099	0.029
blue	0.034	0.002	0.628	0.012	0.151	0.028	0.027

Ground truth	"object"	"rectangle	"red"	"section"	"thin"	"triangle"	"yellow"
green	0.195	0.019	0.000	0.055	0.017	0.079	0.022
red	0.250	0.030	0.946	0.041	0.031	0.072	0.024
yellow	0.210	0.046	0.004	0.063	0.010	0.010	0.740
blue	0.201	0.021	0.006	0.053	0.020	0.084	0.022

Comparative User Study

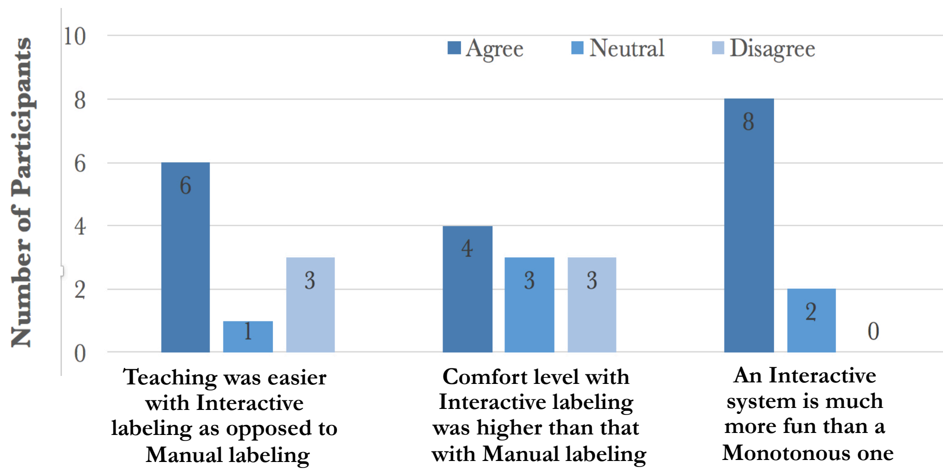
10



- Questionnaire
  - Ease of use
  - Acceptance (willingness to use again)
  - Frustration
- 20% of participants were frustrated
  - Slow response time
  - Anecdotally, system "excessively polite"
  - System requests repetitious

General Concepts

11



Future Work

12

- Incorporate active learning on a more 'real world' problem
  - More complex attributes
  - Non-visual attributes (e.g., spatial relations)
  - More and more complex objects
- Improve metric for classification
  - Verification of confident classifications is a very simple metric
  - Better: entropy of the classifier from the confidence estimate
- Multi-modal interaction
  - Incorporate speech, gesture understanding, …
- Improve interaction!
  - Speed, naturalness, timing and dialog timing

Conclusion

13

- Contribution :
  - Efficient grounded language acquisition with active learning
  - Robot driven interactive labeling system
  - Verification and quality comparison
  - Pilot study of user experiences
- Even simple active learning of language groundings is:
  - *More efficient* than naïve corpus annotation
  - *More pleasant* for users
- More complex approaches are worth pursuing!

Thank you!
Questions?