Zero-shot Learning of Classifiers from Natural Language Quantification

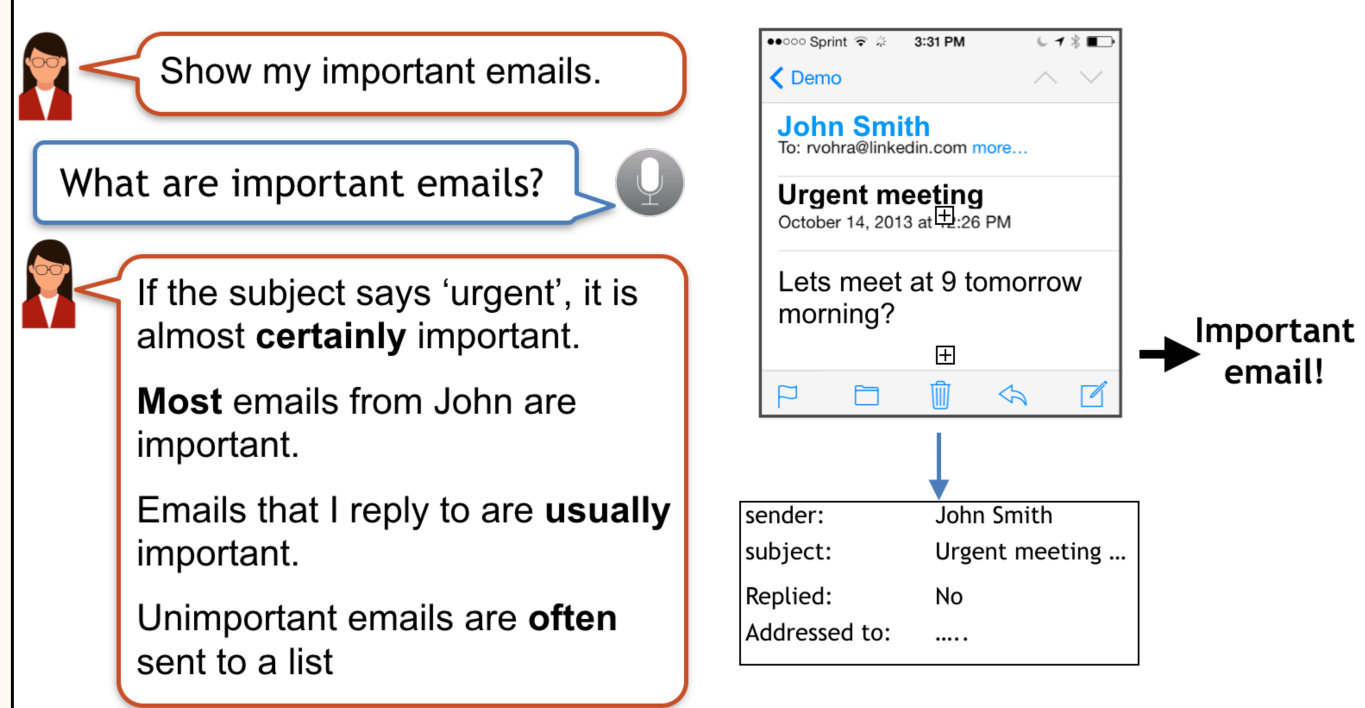
Shashank Srivastava, Igor Labutov, Tom Mitchell

Carnegie Mellon University

MACHINE LEARNING DEPARTMENT

Motivation

Declarative supervision from NL can enable learning with limited or no labeled examples



- Quantifier adjectives and adverbs are explicit denoters of generality
- Use semantics of quantifiers to drive classifier training

Idea

- NL encodes key properties that aid statistical learning

  'Emails that I reply to are usually important'

  1. Features important for a learning problem
     ✓ **x** : repliedTo:true
  2. Class labels
     ✓ **y** : Important
  3. Type of Relationship b/w features and labels
     ✓ P(y|x)
  4. Strength of Relationship
     ✓ Specified by quantifier?

- Convert statements to quantitative assertions

  $$P(\text{label:important} \mid \text{replied:true}) \sim p_{usually}$$

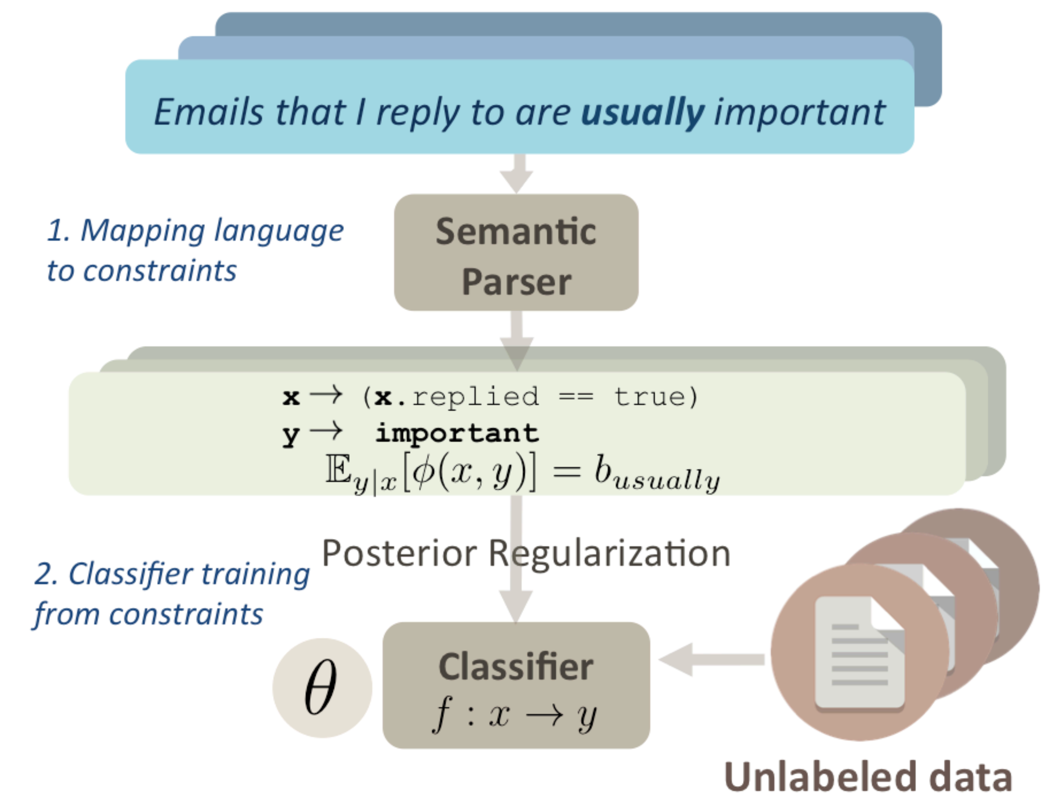
- Simplest model: approximate quantifiers as point probabilities
  - ➢ Pre-register estimates (purely subjective)

Frequency quantifier	Probability value
always , certainly , definitely , all	0.95
usually , normally , generally , likely	0.70
most , majority	0.60
often , half	0.50
sometimes , frequently , some, many	0.30
few , occasionally	0.20
never, rarely	0.05

CONTACT:

Shashank Srivastava
ssrivastava@cmu.edu

Approach



① Mapping Language to Constraints

- Identify key properties to convert each NL explanation of a concept to a quantitative assertion

  $$P(l \mid s) = P(l_{xy} \mid s)\ P(l_{type} \mid l_{xy}, s)\ P(l_{quant} \mid s)$$

- Classifier for constraint types: $P(y\,|\,x)$, $P(x\,|\,y)$ & $P(y)$
  - Based on syntactic and dependency parse features

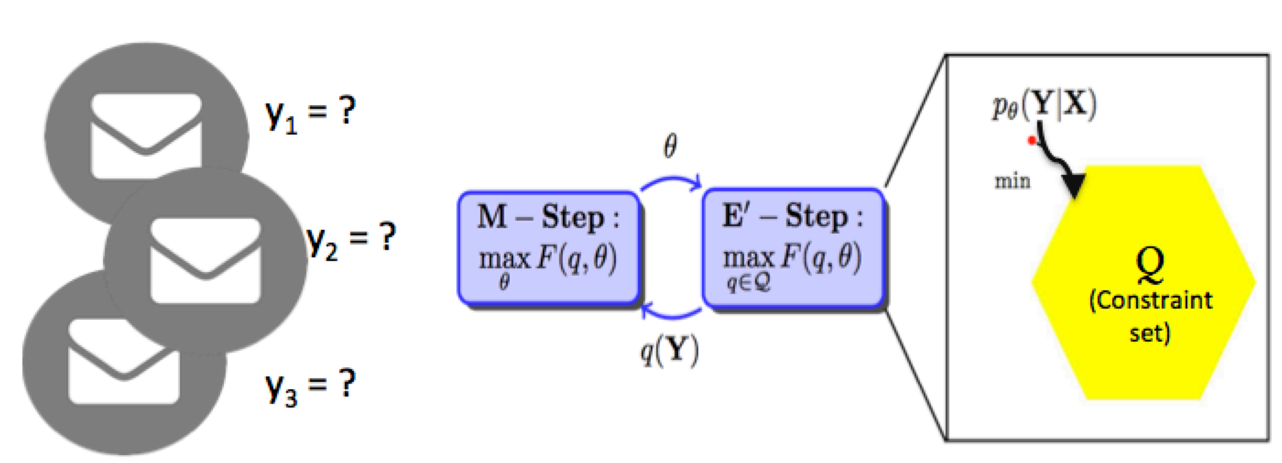
Type	Example
$P(y\,|\,x)$	*Emails that I reply to are usually important*
$P(x\,|\,y)$	*I almost always reply to important emails*
$P(y)$	*About a third of the emails I get are important*

Each assertion acts as a constraint during model training

② Classifier Training via Posterior Regularization

- PR imbues human-provided advice in learned models
- Idea: have predictions from the classifier agree with NL advice on unlabeled data

Class labels for unlabeled examples are latent variables



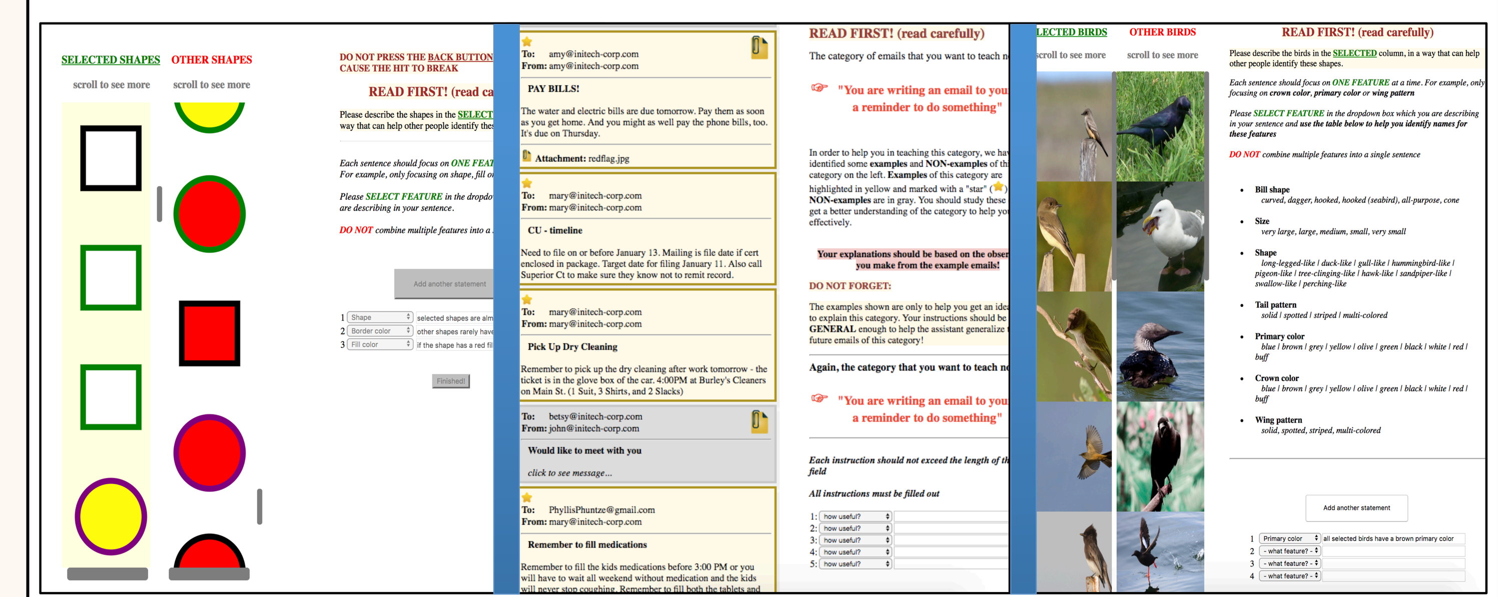
- Modified objective:

$$J_Q(\theta) = \mathcal{L}(\theta) - \min_{q \in Q} KL(q \mid p_\theta(Y|X))$$

Improve data likelihood — Emulate human advice

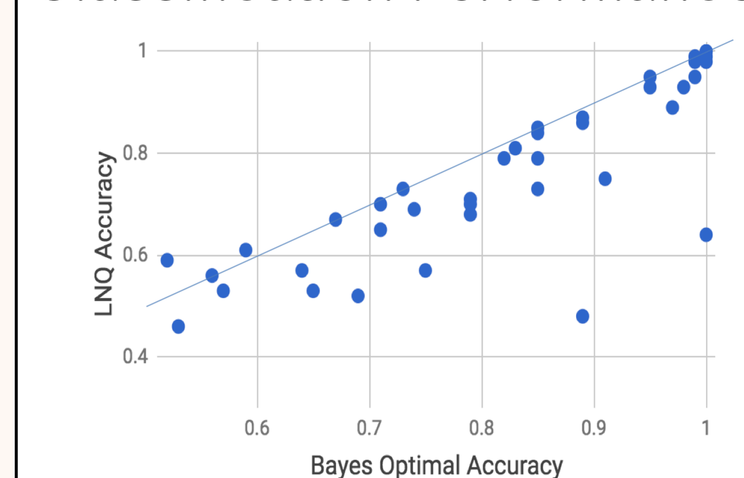
Evaluation

Datasets: Three domains, include synthetic and real concepts



Shape classification — Email categorization — Bird species Identification

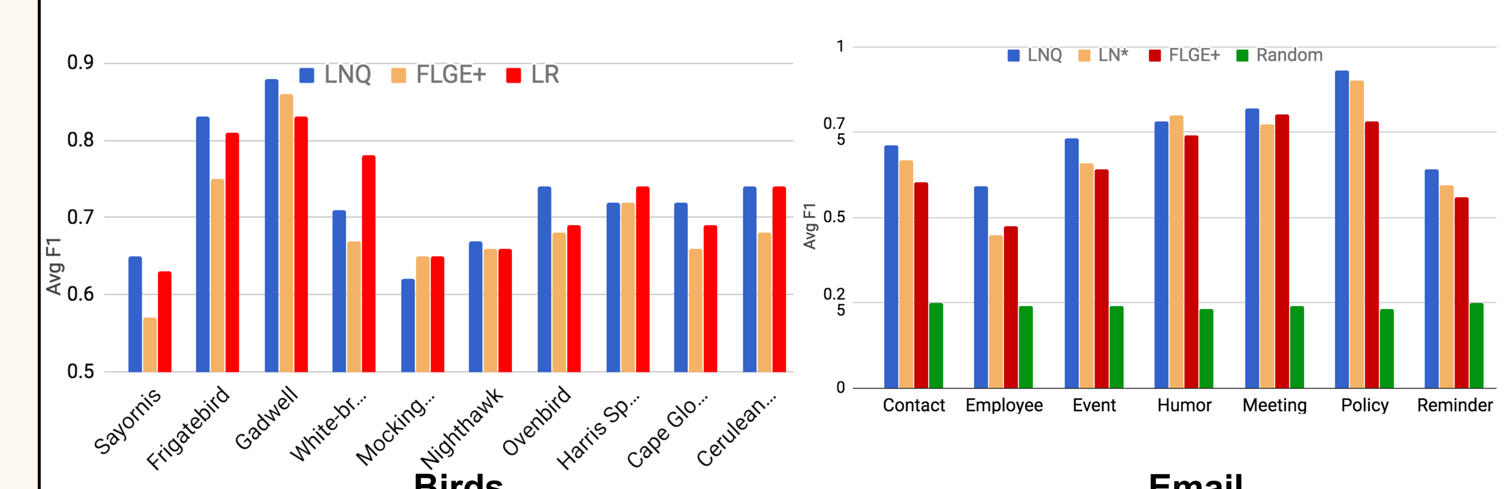
Classification Performance



Approach	Avg Accuracy
Bayes Optimal	0.831
LNQ	0.751
LR (n=10)	0.737
LNQ (coarse quantification)	0.679
LNQ (no quantification)	0.545
Human learner	0.734

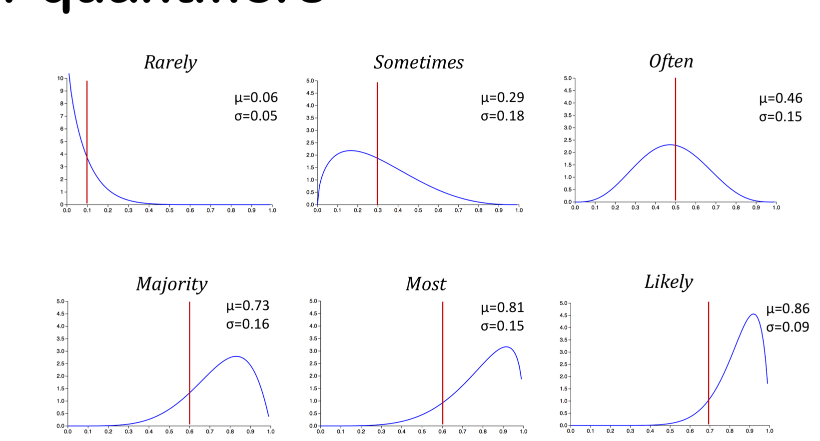
Avg performance on Shape tasks

- **Learning from Natural Quantification (LNQ)** consistently achieves performance comparable with learning from a small number of labeled examples (*LR with n=10*)
- Learning is due to differential associative strength of quantifiers. For *Shape* tasks, LNQ is competitive with humans learning from NL advice



Empirical semantics of quantifiers

- Empirical values may differ significantly from pre-registered beliefs, or show large spreads (not meaningfully modeled as point probability values)



- LNQ is robust to changes in values of probability estimates