

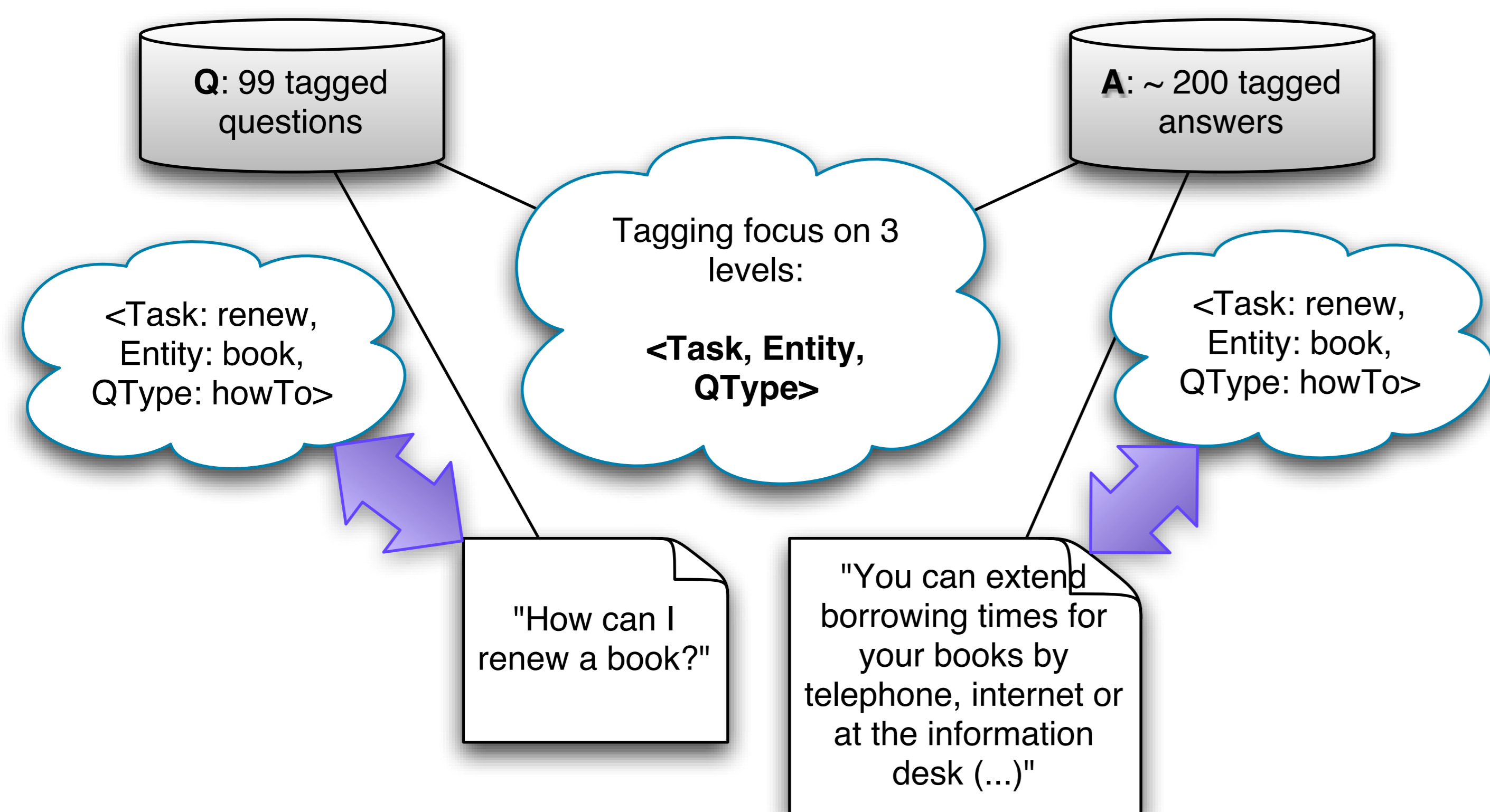
## Introduction and Motivation

- ▶ An Interactive Question Answering (IQA) system has to process users' Follow-Up Questions (FU Q)
- ▶ **Goal:** Predict the focus of a salient FU Q
  - ▶ Can dialogue context help in predicting the FU Q?
  - ▶ Can lexical information help in predicting the FU Q?
- ▶ We introduce a 3-level focus model, and describe a prototype implementation to study focus transitions
- ▶ We use a logistic regression model to learn optimal parameters for the implemented algorithm. The results also indicate properties of salient focus transitions and FU Qs

## Tagging Focus: <Task, Entity, QType>

We represent the (informational) meaning of a user question or a system answer by tagging it with a focus triple consisting of:

- ▶ the **Task** and/or **Entity** that the question/answer is about (= focuses on)
- ▶ the Question Type (**QType**) (e.g., *when*, *howTo*, *yesNo*) that describes the focused target of the question/answer



## A Feature-based Q → A Algorithm

10 features  $x_{1,q,a}, \dots, x_{10,q,a}$  provide information about “fit” of each possible system answer  $a \in A$  to user question  $q$ . Features based on:

- ▶ **1,2,3:** whether surface strings match between  $a$  and  $q$
- ▶ **4:** task/entity structure (i.e., our lexical resource with verbs+arguments)
- ▶ **5,6,7,8:** focus continuity (i.e., dialogue context)
- ▶ **9,10:** task/entity structure + focus continuity

$k$	name	$x_{k,q,a}$	range	description
1	qTypeMatch		0,1	whether QType <sub>a</sub> is matched in $q$
2	taskMatch		0,1	whether Task <sub>a</sub> is matched in $q$
3	lenEntityMatch		n	length of Entity <sub>a</sub> matched in $q$
4	nEntitiesInTask		n	how many of the participant entities of Task <sub>a</sub> (as encoded in our lexical resource = task/entity structure) are matched in $q$
5	taskContinuity		0,1	whether Task <sub>a</sub> is same in $q_{-1}$ and $q$
6	entityContinuity		0,1,2	whether Entity <sub>a</sub> is syn/same in $q_{-1}$ and $q$
7	qTypeContinuity		0,1	whether QType <sub>a</sub> is same in $q_{-1}$ and $q$
8	entityInPrevAnsw		0,1,2	whether Entity <sub>a</sub> is syn/same in $a_{-1}$ and $q$
9	entityInPrevTask		0,1	Focused Task of previous user question has Entity <sub>a</sub> as a participant
10	prevEntityInTask		0,1	Task <sub>a</sub> has focused Entity of previous question as a participant

Table: Features used by the Q → A algorithm

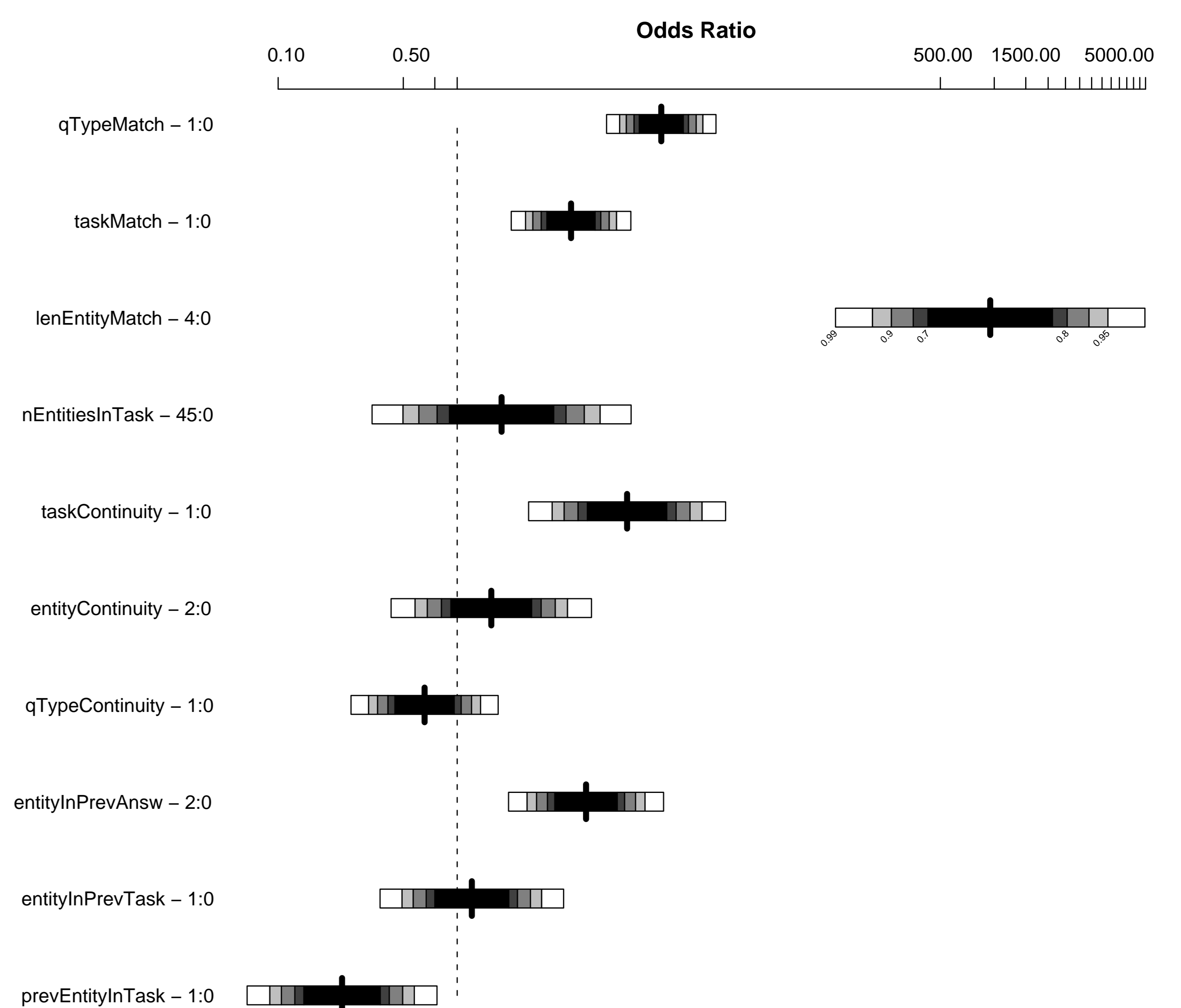
- ▶ **Alg.:** For a new user question  $q$ , iterate through all focus triples  $a \in A$ 
  - ▶ For each combination of  $q$  and  $a$ , evaluate the features  $x_{1,q,a} \dots x_{10,q,a}$
- ▶ **return** highest-scoring answer  $\hat{a}$ :

$$\hat{a} = \arg \max_{a \in A} (\beta_1 x_{1,q,a} + \dots + \beta_{10} x_{10,q,a})$$

## Learning $\beta$ Coefficients: Logistic Regression

- ▶ We needed a principled, data-driven way of setting the “scores”  $\beta_1, \dots, \beta_k$  that each of our 10 features should contribute to the total score for  $a$
- ▶ Logistic regression models describe the relationship between a set of predictors (i.e., our features) and an outcome (i.e., whether a candidate answer is correct for the current  $q$ , thus:  $a = q$ )
- ▶ Training data generated from *all combinations* of  $q \in Q$  and  $a \in A$  by evaluating  $x_{1,q,a}, \dots, x_{10,q,a}$  and using feature values as positive/negative training instance if  $q = a/q \neq a$
- ▶ We use the logit  $\beta$  coefficients that a logistic regression model estimates for each predictor (from training data, using maximum likelihood estimation) as our empirically learned scores

## Results



**Interpretation** of significant model coefficients/features shown above:

- ▶ all **3 surface-based features** are good predictors of correct answer
- ▶ the history-based feature **taskContinuity** also indicates correct answers: “topic exploration”, with the FU Q asking about a different aspect of the previously focused task
- ▶ **entityInPrevAnsw** indicates correct answer; shows influence of the system answer in shaping the IQA dialogue
- ▶ **prevEntityInTask** actually seems to indicate a *false* answer

**Evaluation in terms of answer accuracy (%)**, done on question collection Q:

- ▶ With manually set feature scores. 24 out of 78 questions answered with correct focus triple ( $a = q$ ) from total of > 200 answer candidates. **Accuracy: 30.8%**
- ▶ With learned feature scores.  $\beta$  coefficients from logistic regression model with all 6 significant predictors shown above. 47 out of 78 test questions correct. **Accuracy: 60.3%**

## Conclusions

We presented a Q → A algorithm implementing our model of focus for IQA, and demonstrated how this algorithm’s accuracy improves with parameters learned from data. To predict the focus of a salient FU Q, we have shown that an IQA system needs additional knowledge about:

- ▶ dialogue context: the previously focused task, and the entities mentioned in the previous system answer
- ▶ (possibly also: lexical information, as in prevEntityInTask)