Learning to Ask for Conversational Machine Learning

Shashank Srivastava[1*] **Igor Labutov**[2] **Tom Mitchell**[3]
[1]UNC Chapel Hill, [2]LAER AI
[3]Machine Learning Department, Carnegie Mellon University
ssrivastava@cs.unc.edu, igor.labutov@laer.ai, tom.mitchell@cmu.edu

Abstract

Natural language has recently been increasingly explored as a medium of supervision for training machine learning models. Here, we explore learning classification tasks using language in a conversational setting – where the automated learner does not simply receive language input from a teacher, but can proactively engage the teacher by asking template-based questions. We experiment with a reinforcement learning framework, where the learner's actions correspond to question types and the reward for asking a question is based on how the teacher's response changes performance of the resulting machine learning model on the learning task. In this framework, learning good question-asking strategies corresponds to asking sequences of questions that maximize the cumulative (discounted) reward, and hence quickly lead to effective classifiers. Empirical analysis shows that learned question-asking strategies can expedite classifier training by asking appropriate questions at different points in the learning process. The approach allows learning using a blend of strategies, including learning from observations, explanations and clarifications.

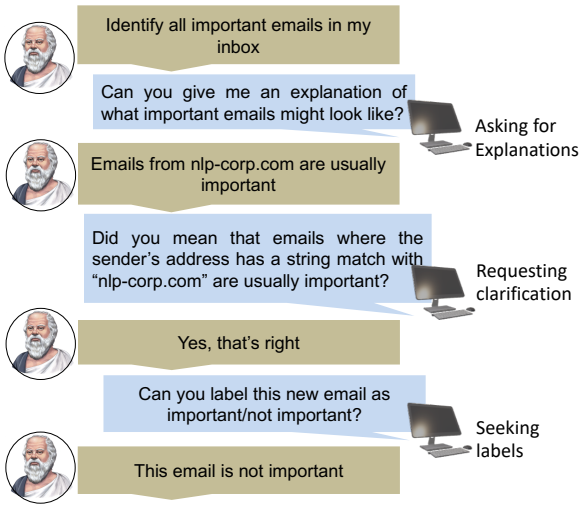

Figure 1: Question-Answer dialog can enable learning from a mix of strategies, including label observations (traditional supervised learning), explanations and clarifications (to overcome parsing limitations). The output from the teacher-learner interaction is a classification model (here, for important emails). We present a framework that (a) enables learning classifiers from a mix of such supervision; (b) learns to ask appropriate sequences of questions to accelerate this.

1 Introduction

The ability to learn new tasks and behaviors from language is characteristic of human intelligence. In recent years, the fields of machine learning and NLP have seen an renewed interest in incorporating natural language supervision in models of machine intelligence (Narasimhan et al., 2015; Elhoseiny et al., 2013; Goldwasser and Roth, 2014; Andreas et al., 2018; Fried et al., 2018; Wang et al., 2016). In particular, methods such as BabbleLabble (Hancock et al., 2018) and LNL (Srivastava et al., 2017) show progress towards realistic applications of supervised learning from language on tasks such as information extraction and email categorization. However, until now, such methods have been limited in two ways.

First, despite a body of work on leveraging language for tasks involving human robot interaction (She and Chai, 2017; Cakmak and Thomaz, 2012; Krishnamurthy and Kollar, 2013) and interactive learning in non-linguistic settings (see Section 2), existing approaches for training machine learning models from language are largely non-interactive, i.e. the learner agent receives statically collected text-based advice from a teacher as input, but does not directly engage with the

* Work done while the first and second authors were at CMU

Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing, pages 4164–4174, Hong Kong, China, November 3–7, 2019.

teacher.[1] In comparison, when humans learn, they do not rely only on passively receiving instruction from a teacher. Rather, the interaction takes the form of a mixed-initiative dialog, where they ask questions and proactively seek clarifications to simplify learning. These questions can generalize learning to novel situations, explore hypotheses, or fill information gaps. The ability to ask questions can, thus, fundamentally facilitate learning.

Second, existing approaches have focused on using language either as a standalone replacement for labeled data (Hancock et al., 2018), or to drive learning such as through specifying features for learning tasks (Eisenstein et al., 2009). In contrast, many realistic scenarios of learning from language would involve not learning from language alone, but learning from a mix of supervision, including both traditional labeled data, and natural language advice. Thus, automated learners should be capable of learning from a blend of observations, explanations and clarification.

In this work, we introduce a framework for learning from language in a conversational setting (*LiD*, for *Learning with Interactive Dialog*), which is a step towards alleviating these shortcomings. Language provides a natural medium for conversational interactions between a learner and a teacher, specifically in the form of question-answer dialog. The premise driving our work is that the ability to ask questions can be leveraged by an automated learner to accelerate its learning. We explore a data-driven approach for learning effective question-asking strategies in the specific context of learning classification tasks. The signal for learning to ask questions is grounded in the learning task itself. i.e., the value of a question is evaluated in utilitarian terms of how it affects performance on a downstream classification task. This follows a Wittgensteinian view of language as a cooperative game (Wittgenstein, 1953) between agents (here, the teacher and a learner) with a shared goal (here, building an effective classifier). While the space of questions that an interactive learner can ask can be vast in general, here we specifically focus on leveraging interactivity for three specific aspects (highlighted in Figure 1):

1. Seeking labels for specific examples.
2. Asking for explanations of a concept.
3. Requesting clarifications about explanations.

As illustrated in Figure 1, these dimensions can facilitate multiple aspects of the learning process: including learning from labeled examples (similar to traditional supervised learning), learning from natural language explanations (similar to recent work on learning from explanations) and alleviating limitations in the learner's semantic parsing abilities (in vein with work such as Labutov et al. (2018)). Learning systems that reify these abilities can enable users to interactively teach new concepts using a blend of traditional and language-based supervision. Our contributions are:

- A reinforcement learning formulation to guide question-asking strategies for learning from language.
- A method for interactively training classifiers using a mix of labeled data, natural language explanations and clarifications. Our exploration highlights some of the challenges involved in interactive learning from language.

2 Challenges in Relation to Previous Work

From the perspective of traditional supervised learning, the problem of asking questions can be seen as cognate with active learning. Methods in active learning have explored various criteria for choosing which of a set of unlabeled examples to label next while training supervised machine learning models (Settles, 2012; Collins et al., 2008). This can be seen as asking a specific kind of question (as illustrated in Figure 1). Learning to ask questions generalizes active learning in multiple ways by possibly soliciting a wider range of data measurements. These include feature labels ('Are emails with subject "urgent" usually important?'), label proportions ('Around what fraction of emails are important?'), constraints on model expectations ('Are you more likely to reply to important emails?'), etc. Approaches such as Srivastava et al. (2018) map such language to data measurements that computational models can reason over.[2] Statistical frameworks such as Generalized Expectation (Druck et al., 2008), Posterior Regularization (Ganchev et al., 2010) and Bayesian

[1] Zhang et al. (2018) diverge from prior work in this respect, and model language games between teachers and learners. However, their learning tasks are toylike, and the method does not generalize to realistic scenarios.

[2] For example, a statement such as 'Emails from my boss are usually important' may be mapped to a data measurement of form $P(y = important|sender = boss) \approx P_{usually}$.

Measurements (Liang et al., 2009) then allow for model training from a broad range of such data measurements in conjunction with unlabeled data, rather than using labeled examples. Other recent approaches such as Huang et al. (2015) and Siddiquie and Gupta (2010) have expanded pool-based active learning to learning from multiple types of queries, especially in the context of multi-label and multi-class learning. Similarly, Parikh and Grauman (2011) explore feature space construction for visual tasks in an interactive setting. Although in principle soliciting different types of data measurements can help learning, each type requires its own interface. The advantage of using natural language as a medium is that it allows us to unify the different modes of supervision into a single, familiar user interface. However, using natural language as a medium of supervision comes with its own set of challenges, as we discuss next.

2.1 Dependence on Language Interpreter

Since both generation and transmission of language advice can be noisy, the optimal question asking strategy may depend not only on the information content of data measurements, but also factors such as the quality of the learner's semantic parsing model and the teacher's skill.[3] To explain, while useful from an information theoretic sense, a teacher's explanations may be too complex to handle for the learning agent's parser, in which case it might be preferable to stick to asking about the teacher about instance labels (which would require minimal parsing). Thus, question-asking strategies need to be sensitive to the learner's own semantic parsing ability, which may also change during the course of interactions with users.

2.2 Context Dependence

Rather than learning a static criterion for choosing what question to ask (as in active learning), our focus is on asking questions in conversational settings, which are inherently dynamic processes. To explain, asking a teacher to rephrase an explanation only makes sense in specific contexts (when the interpretation of something said previously is unclear). Further, the question to ask can depend on factors such as the task domain, supervision previously received, etc. These factors motivated our choice of a reinforcement learning approach for learning question-asking strategies.

Relation to Question Generation approaches: The problem of learning to ask questions has previously been explored by several approaches. Vanderwende (2008) and Olney et al. (2012) explore generating reading comprehension questions conditioned on a given text. More recently, Romeo et al. (2016) and Rao and Daumé III (2018) present neural network models that rank questions in community QA forums, whereas Misra et al. (2018) generate questions for visual scene understanding. Our framework significantly differs from these in its sequential framework, and that the questions to be asked are grounded in quantitative performance on a downstream task.

3 Approach

In this section, we describe our framework for interactive learning from question-answer dialog. We first describe an approach to learn classifiers using a mix of explanations and labeled examples in Section 3.1. This is a preliminary towards question-asking strategies that subsume active learning as well as language advice; and constitutes a subroutine that is repeatedly invoked in our approach. Section 3.2 describes our reinforcement learning formulation for learning question-asking strategies in simulated conversational settings. The space of actions consists of a vocabulary of question types that a learner can ask, and reward is based on improvements in the classification model that the teacher's response to an asked question leads to.

3.1 Learning classifiers from a mix of observations and explanations

We base our learning framework on previous work by Srivastava et al. (2018), who train loglinear classifiers (with parameters θ) using natural language explanations of the individual classes and unlabeled data. Further, they use the semantics of linguistic quantifiers (such as 'usually','always', etc.) as priors in a Posterior Regularization objective to drive the model training. In particular, their training objective takes the following form:

$$J_Q(\theta) = \underbrace{\mathcal{L}(\theta)}_{\text{Explain data}} \underbrace{- \min_{q \in Q} KL(q \mid p_\theta(Y|X))}_{\text{Emulate human advice}} \quad (1)$$

[3] In this work, we are not interested in learning semantic parsing models. We presume the existence of pretrained semantic parsers for learning agents. Our focus is rather on whether some question asking strategies may be more effective than others for a learning agent with given capabilities.

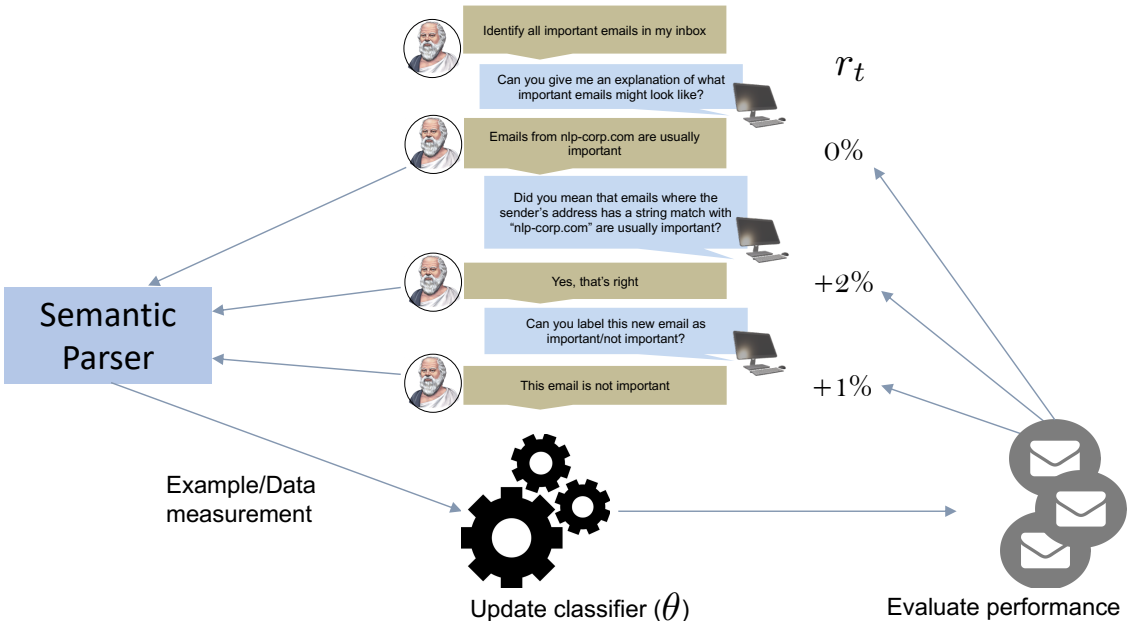

Figure 2: We assume that the dialog between the learner and the teacher is in the form of turn-wise conversations – consisting of a sequence of questions asked by the learner, and the teacher's responses to those questions. At each step in this process, the teacher's response is parsed by the learner (using a pre-trained semantic parser), and can be incorporated into the learner's concept model as either a labeled example or a data measurement (the learner can also choose to seek a clarification). A reward (denoted by r_t) can be computed at each step, which denotes the marginal change in classification performance on a held-out set of examples due to the last response. In this framework, learning good question-asking strategies corresponds to asking sequences of questions that maximize the cumulative (discounted) reward, and hence quickly lead to effective concept models. The framework also allows for asking sequences of multiple questions before seeing a major jump in model performance.

The objective reflects a tension between explaining the unlabeled data (likelihood term) and emulating the natural language advice provided by a teacher. The KL divergence represents difference between predictions from the trained model on unlabeled data $p_\theta(Y|X)$ and language advice (each explanation is incorporated as a data measurement; the conjunction of these defines the 'valid set' of posterior distributions Q that perfectly concur with the natural language advice). The second term essentially computes the minimum distance between the model posterior and the set Q.

Here, we show that we can naturally extend this approach to learn classifiers from a mix of both labeled and unlabeled data, and natural language explanations. To do this, we simply append a log-likelihood term for the labeled examples to the objective in Equation 1. The updated objective is:

$$J_Q(\theta) = \mathcal{L}_{labeled}(\theta) + \mu\Big(\mathcal{L}_{unlabeled}(\theta) - \min_{q \in Q} KL(q \mid p_\theta(Y|X))\Big) \quad (2)$$

Here, $\mathcal{L}_{labeled}(\theta)$ denotes the log-likelihood term for a set of $n_{labeled}$ labeled examples $\mathcal{X}_{labeled} = \{(x_k, y_k)\}_{labeled}^{n}$ (normalized by $n_{labeled}$), whereas the other two terms are as before: $\mathcal{L}_{unlabeled}(\theta)$ denoting log-likelihood over a set of $n_{unlabeled}$ unlabeled examples, and a posterior regularizer term (KL-divergence) penalizing violations of the parse natural language advice. In the E-step of the Posterior Regularization training (Ganchev et al., 2010), the computation of the posterior regularizer remains unchanged. However, the M-step is modified so that the classifier parameters θ are learned using both the inferred labels for the unlabeled data, and provided labels for the labeled examples.

In Equation 2, $\mu > 0$ determines the relative weights of provided example labels and natural language advice in the optimization objective and is a hyper-parameter for the method. In learning scenarios where there is little labeled data, we would like to rely primarily on constraints specified from natural language explanations, and unlabeled data. On the other hand, in scenarios where there is a lot of labeled data available enabling robust inductive inference, we would like

to primarily rely on it rather than explanations.[4] While setting up the optimization problem, the value of μ can be adapted to reflect this intuition. In our experiments, we found setting $\mu = 1/\max(n_{labeled}, 1)$ to work well across settings.

3.2 RL formulation for learning to ask

Figure 2 illustrates our framework for learning classification tasks in a question-answer dialog setting. We assume the presence of a teacher to answer questions posed by the learner. We restrict the structure of dialog to a sequence of questions $(q_1 \dots q_T)$ asked by the learner, and the teacher's responses $(u_1 \dots u_T)$ to them. We further assume the presence of a held-out set of labeled examples of the concept, which can be used to evaluate the learner's classification performance as the dialog progresses. At each step t, the learner's action a_t consists of choosing a question to ask the teacher. The teacher's response to the learner's question is parsed (in the form of a labeled example, or a data measurement), which is then incorporated into the learner's concept model (by retraining with the additional labeled example or the new data measurement). The classification performance, c_t, of the updated model is evaluated on the held-out set. The change in classification performance from the previous step, r_t approximates the marginal value of the question in learning the task, and constitutes the learner's reward at that step.

Our approach for learning question asking strategies models the dialog as a simple Markov Decision Process. Since our state and action spaces are discrete (as described in the following sections), we can use a table-based SARSA-learning procedure (which allows for on-policy learning over Q-learning) to estimate the state-action values $Q(s, a)$ of different question types in different contexts. We next describe the state-space, actions and rewards, and the learning procedure.

3.2.1 Action Space

As mentioned earlier, there can be a multitude of questions that a learner can ask a teacher. Here, we are interested in exploring three specific types of questions that are specially germane for facilitating learning from a mix of labeled examples and explanations. These consist of the following:

1. **Seeking labels for specific examples:** This is similar to traditional active learning. In particular, we can have a different action corresponding to every active learning criterion, which chooses which example to label next. In our experiments, we use two active learning techniques (with a corresponding action for each):
   - *Random:* Ask for class label for a randomly chosen unlabeled instance.
   - *Maximum Uncertainty:* Ask for class label for the instance in the data for which the current concept model is most uncertain (highest entropy). If there are multiple such instances, randomly pick one among them.
2. **Asking for an explanation for the concept:** This action seeks out from the teacher a short natural language explanation of the concept. This is then incorporated in the concept model as a quantitative constraint. In general, this can encompass several types of questions:
   - Asking for probability estimates about specific labels and features. e.g., 'How often are emails about meetings important?'
   - Asking for discriminative features for particular concept labels. e.g., 'Can you think of a feature that if present always denotes that an email is important?'
   - Asking about class probabilities. e.g., 'Around what fraction of emails in your inbox are important?'

   In principle, each of the above provide admissible constraints which the classifier training procedure (from Section 3.1) can handle. However, to simplify analysis, in our experiments, we conflate these actions into one category, and ask for general explanations of the form '*Can you give me an explanation of the concept?*', which could return a variety of data measurements (subsuming $P(y)$,$P(y|x)$ and $P(x|y)$, previously explored in Srivastava et al. (2018)).
3. **Requesting clarification about the previous explanation:** This action asks for a clarification about the interpretation of a previous explanation (which can be helpful in cases when the learner is uncertain about the interpretation). For this, the learner verifies if the interpretation of the previous explanation (using

[4] In the asymptotic case of infinite data with labels, an inductively learned Bayes Classifier would be optimal.

the learner's semantic parser) was correct or not. To do this, we generate a question of the form '*Did you mean ...*' using a synchronous grammar which deterministically maps logical forms to natural language descriptions (see Figure 1 for an example). The teacher responds with yes, if the parsed logical form matches the gold annotated logical form, and with no otherwise. In case the teacher responds with no, the current explanation is discarded (not used in model training), and the learner moves ahead to ask for a new label or explanation.

Simulating Interactions: We note that each of the question types described above – (1) asking for labels for examples, (2) asking for concept explanations, and (3) verifying interpretations of language explanations – can be simulated with corresponding statically collected data – consisting of (1) labeled examples for classification, (2) natural language explanations of classes, and (3) annotations of those explanations with logical forms. This has a significant implication: rather than relying on questioning human users in real-time, we can simulate the conversational exchange by asking questions to an oracle, which has access to previously pre-collected data of the above-described form for each classification task. While this is a coarse approximation of actual dialog between an automated learner and human teachers, it can serve as a useful proof-of-concept, and allows for quick experimentation. We rely on this simulated setting for learning policies for question asking.

3.2.2 Rewards

The reward, r_t, evaluates the change in classification performance due to an asked question at each step t of the dialog. The performance of the classification model is evaluated on a held out set of $n_{heldout} = 50$ labeled examples for each learning task. In our experiments, we use the model's F1 score as the metric for classification performance, c_t. We define the reward as the absolute change in model performance from the previous step:

$$r_t = c_t - c_{t-1} \tag{3}$$

3.2.3 State-space

Next, we describe the featurized state space for our reinforcement learning formulation. The best question to ask at a particular point can likely depend on the state of the conversation. This could include factors such as the pedagogical phase in the learning process (exploratory vs confirmatory), previous questions asked, etc. Thus, defining a rich enough state space is an important consideration for a formulation of conversational learning. In our treatment, we assume a discrete state-space, which is defined as the cross product of the following (also discrete) features.

- **Curricular stage in the Learning process:** We use a discrete variable to model the curricular stage of the learner, approximating it by the number of steps (questions previously asked) in the interaction at any point. We cluster the number of steps in the following five bins of values: BEGINNER (0 steps), NOVICE (1-5 steps), INTERMEDIATE (6-10 steps), ADVANCED (11-15 steps) and MATURE ($>$ 15 steps).[5]
- **Reward in the previous state:** We discretize the value of reward as belonging to one of four ranges (abstractly named GOOD, INCREASING, FLAT, and DECREASING), with thresholds chosen to correspond to the inter-quartile ranges for the value of rewards observed in evaluating a random policy.
- **Velocity of reward:** This is a ternary variable indicating whether the value of the discrete variable for the reward (above) is BETTER,WORSE or the SAME than the previous step.
- **Type of the previous two actions:** As mentioned in the description of the action space.
- **Domain of learning task:** Indicates the domain of the current classification task. We use datasets corresponding to three domains: EMAIL CATEGORIZATION, SHAPE CLASSIFICATION and BIRD SPECIES IDENTIFICATION; hence this variable can take these three values.
- **Confidence of previous parse:** We model the learner's confidence in parsing the response u_t from a teacher as the ratio of the probability of the highest probability (predicted) logical form from the learner's semantic parser and the next best logical form. We discretize this ratio into three values, corresponding to the upper (HIGH), lower (LOW) and middle two inter-quartile ranges (MEDIUM) for the value of the ratio over all explanations in our data.

[5]These threshold values were heuristically chosen based on the observation that for many datasets that we experiment with, classifier performance roughly begins to taper at around 20 training examples.

While our state space captures several facets, it does not model some other important factors:

- **Teacher behavior:** Whether the teacher provides correct information, and uses easily interpretable language.
- **Task difficulty:** This refers to how expressible a classification task is using language explanations. For example, some concept maybe significantly easier to explain using language than others, depending on the logical language available to the semantic parser. For example, it maybe impossible to explain digit recognition using pixel level features using language.

3.3 Model training

Since the action and state-spaces are discrete and not prohibitively large, we use the on-policy SARSA learning algorithm for policy control (Rummery and Niranjan, 1994), where we represent the state-action Q-values for pairs of states s and actions a as a table. We use an ϵ-greedy strategy ($\epsilon = 0.20$). i.e., the strategy balances between exploitation and exploration by picking the next action to be the estimated optimal one (having the maximum estimated Q-value for a state) with probability $1 - \epsilon$, and choosing the next action randomly with a probability of ϵ. The initial policy is defined by uniform randomly initializing $Q(s, a)$ values between 0 and 1.

4 Data

Our empirical analysis uses existing datasets for learning classifiers from language. These are datasets for email categorization from natural language explanations from Srivastava et al. (2017); and bird species classification and synthetic shape classification tasks from Srivastava et al. (2018). In all, thesee consist of 67 classification tasks belonging to these three domains.[6] For each task, the corresponding data consisting of natural language explanations of classes as well as annotated logical forms for these explanations are available. For each task, we hold out a random sample of 50 examples for evaluating classifier performance. The rest of the examples (ranging between 50 and 100 for individual tasks) are considered as unlabeled data at the start of each interaction. At each step in a simulated interaction, either (1) the label for a new example from this set is revealed by the teacher, (2) a new data measurement is provided as a natural language explanation by the teacher, or (3) the learner seeks a clarification about the parse of the previous explanation. Thus, from the learner's perspective, over time the size of the unlabeled data decreases, and the labeled examples increase (see Equation 2).

[6]The complexity of these problems varies considerably, ranging from marginally above chance to perfectly learnable.

5 Experiments

In this section, we evaluate *LiD*'s performance for three domains of classification tasks. As previously mentioned, for policy learning we simulate conversational interactions between teachers and learners by asking questions to a oracle, which has access to previously collected data about each classification task. One limitation of learning question-asking strategies from simulated interactions is that for some classification tasks, we may run out of explanations in the course of model training (since the number of explanations of a concept are limited). In these cases, we end the interaction as soon as all explanations are already provided to the learner.

5.1 Learned vs Random policy evaluation

First, we compare learned policies for question asking with a naive policy that randomly takes a new action at each step in the learning process.

Figure 3 shows averaged cumulative reward for question asking strategies on 20 unseen classification tasks, after SARSA learning for 10 epochs on the remaining 37 classification tasks. In the figure, the x-axis corresponds to the number of steps in a dialog, and the y-axis denotes the cumulative reward (averaged over 20 tasks) for a learned policy vs a random policy. We note that policy learning leads to consistently superior performance (on average, *LiD* achieves any given level of classification performance in fewer steps), which unambiguously indicates value in asking the right sequences of questions. We observed that the trend was also similar in most individual learning tasks. For example, the cumulative reward after 10 steps was higher for the learned policy than the random policy for 18 of the 20 learning tasks. The difference in performance was statistically significant at $p < 0.05$ using a signed permutation test. We characterize some learned behaviors that drive this improved performance in Section 5.4.

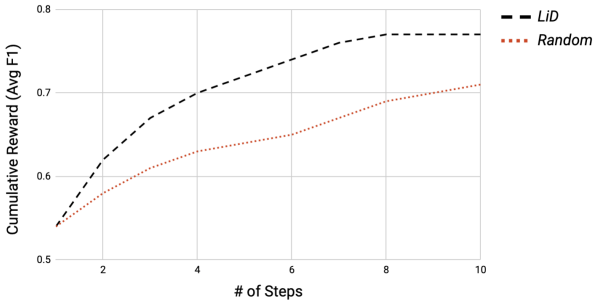

Figure 3: Cumulative Reward (averaged over 20 tasks) for interactive concept learning for learned question asking policy vs random policy.

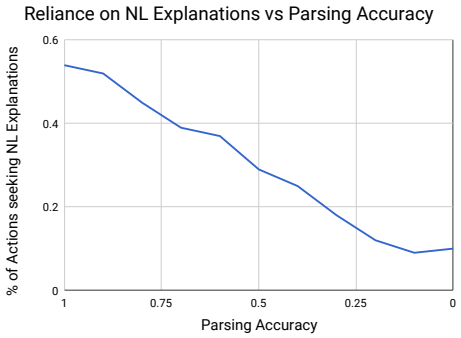

Figure 4: Fraction of actions seeking Natural Language Explanations vs competence of the learner's semantic parser.

5.2 Reliance on NL vs Parsing accuracy

Intuitively, semantic parsing competence of a learner should be a significant consideration in whether it should rely on explanations. To test this, we simulate scenarios of learners with different levels of semantic parsing ability by choosing the true logical form for any explanation with the corresponding probability, and choosing an alternative logical form from the remaining candidates in the k-best list from the semantic parser otherwise. Figure 4 depicts the effect of parsing competence on learned question-asking strategies. The empirical behavior largely corroborates our expectation, as the learned strategies increasingly avoid seeking natural language explanations of concepts as the parsing performance worsens. In the base case where the learner has no parsing competence, the model learns to exclusively ask for labeled examples only (In the figure, the fraction is seen to converge close to 0.1 rather than 0 due to the ϵ-greedy nature of the policy).

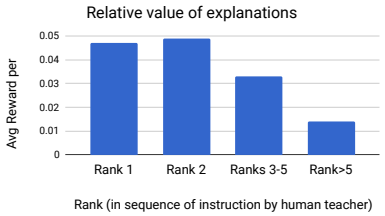

Figure 5: Average increase in classification performance from explanations of different positional orders.

5.3 Differential value of explanations

A notable issue in learning from explanations, which we do not model here is that a teacher's multiple explanations of a concept can have a large variance in their utility to a learner. In particular, we might expect that teachers would be more likely to provide the most useful explanations first, and minor explanations subsequently. From an ablation study, we observe that this is indeed a valid concern. Figure 5 shows the average marginal increase in classification performance over 50 visual shape classification tasks from explanations with different rank (based on the actual order of providing from human users). This indicates that explanations from teachers provided later contribute significantly less towards classification performance.

5.4 Examples of Learned behavior

From a qualitative analysis, learned policies are seen to be intuitive and interpretable. In particular, the policies overwhelmingly seek clarifications when confidence in the parser is low. On the other hand, there are strong inclinations to continue using an action type as long as it yields high returns. Interestingly, the optimal policy differs significantly in behavior across domains. As example, learned policies rely nearly twice as much on explanations for bird species identification as for email classification tasks. The probable reason is that parsing is harder for email explanations, as features in this domain are often compositional.

5.5 User study

We perform a small user-study to also evaluate performance of the learned policies on an email categorization task with actual human teachers. We still train the question-asking strategy using the simulated teacher framework (since learning the policy from crowdsourced human users would be expensive and slow). 20 users were asked to in-

	Avg. Reward	Natural	Avg. Rew (simulated)
LiD	0.524	3.2	0.607
Random	0.493	2.9	0.551

Table 1: Human teacher evaluation for learned and random question asking policy.

teract with the learned *LiD* policy to teach a chosen email-classification task. For each task, the system asked a sequence of 10 questions, and the human teacher's responses were incorporated into the system to update the classification model. The users were also asked to teach another task with questions asked through a random policy. Table 1 shows the average cumulative reward for humans interacting with *LiD* vs a random policy for this experiment. We note that *LiD* leads to better performance on average. This trend is the same as in the simulated analysis, although we note that the learning is slower with real teachers than in the simulated setting on the same tasks, and the gain in performance is substantially smaller. A contributing reason for this is likely annotator bias (Geva et al., 2019), since in the simulated testing scenarios, the teacher's explanations can often likely come from a small set of turkers whose language explanations for teaching other tasks were used for training the learner's semantic parsing model. We note that the learned policy was rated by human users as more natural than a random policy on a Likert scale (with range 1-5).

6 Conclusion

In this paper, we have provided a reinforcement learning formulation for learning to ask questions for interactive training of machine learning models. This framework is attractive in grounding the value of questions asked in a measurable downstream task. Further, change in model performance is a natural reward to drive this learning. While this provides a conceptually useful framework for framing question generation, in its current form the approach makes simplistic assumptions on the types of questions that can be asked, as well as on the structure of the dialog between the teacher and the learner. While the system outperforms a random policy on learning classification tasks, the dialog looks contrived from a human perspective. An interesting direction could be to pair the framework with neural text generation methods to model fine-grained question types, and generate more natural-looking interactions through dialog. An important scientific question is to characterize learning tasks for which learning from language is likely to outperform pure inductive learning. Future work can also extend the approach to other supervised learning tasks, as well as bootstrap from natural dialog data.

References

Jacob Andreas, Dan Klein, and Sergey Levine. 2018. Learning with latent language. In *Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long Papers)*, pages 2166–2179. Association for Computational Linguistics.

Maya Cakmak and Andrea L. Thomaz. 2012. Designing robot learners that ask good questions. In *Proceedings of the Seventh Annual ACM/IEEE International Conference on Human-Robot Interaction*, HRI '12, pages 17–24, New York, NY, USA. ACM.

Brendan Collins, Jia Deng, Kai Li, and Li Fei-Fei. 2008. Towards scalable dataset construction: An active learning approach. In *European conference on computer vision*, pages 86–98. Springer.

Gregory Druck, Gideon Mann, and Andrew McCallum. 2008. Learning from labeled features using generalized expectation criteria. In *Proceedings of the 31st annual international ACM SIGIR conference on Research and development in information retrieval*, pages 595–602. ACM.

Jacob Eisenstein, James Clarke, Dan Goldwasser, and Dan Roth. 2009. Reading to learn: Constructing features from semantic abstracts. In *Proceedings of the 2009 Conference on Empirical Methods in Natural Language Processing: Volume 2-Volume 2*, pages 958–967. Association for Computational Linguistics.

Mohamed Elhoseiny, Babak Saleh, and Ahmed Elgammal. 2013. Write a classifier: Zero-shot learning using purely textual descriptions. In *The IEEE International Conference on Computer Vision (ICCV)*.

Daniel Fried, Ronghang Hu, Volkan Cirik, Anna Rohrbach, Jacob Andreas, Louis-Philippe Morency, Taylor Berg-Kirkpatrick, Kate Saenko, Dan Klein, and Trevor Darrell. 2018. Speaker-follower models for vision-and-language navigation. *CoRR*, abs/1806.02724.

Kuzman Ganchev, João Graça, Jennifer Gillenwater, and Ben Taskar. 2010. Posterior regularization for structured latent variable models. *The Journal of Machine Learning Research*, 11:2001–2049.

Mor Geva, Yoav Goldberg, and Jonathan Berant. 2019. Are we modeling the task or the annotator? an investigation of annotator bias in natural language understanding datasets. *arXiv preprint arXiv:1908.07898*.

Dan Goldwasser and Dan Roth. 2014. Learning from natural instructions. *Mach. Learn.*, 94(2):205–232.

Braden Hancock, Paroma Varma, Stephanie Wang, Martin Bringmann, Percy Liang, and Christopher Ré. 2018. Training classifiers with natural language explanations. In *Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 1884–1895. Association for Computational Linguistics.

Sheng-Jun Huang, Songcan Chen, and Zhi-Hua Zhou. 2015. Multi-label active learning: Query type matters. In *Proceedings of the 24th International Conference on Artificial Intelligence*, IJCAI'15, pages 946–952. AAAI Press.

Jayant Krishnamurthy and Thomas Kollar. 2013. Jointly learning to parse and perceive: Connecting natural language to the physical world. *Transactions of Association for Computational Linguistics*.

Igor Labutov, Bishan Yang, and Tom Mitchell. 2018. Learning to learn semantic parsers from natural language supervision. In *Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing*, pages 1676–1690.

Percy Liang, Michael I Jordan, and Dan Klein. 2009. Learning from measurements in exponential families. In *Proceedings of the 26th annual international conference on machine learning*, pages 641–648. ACM.

Ishan Misra, Ross Girshick, Rob Fergus, Martial Hebert, Abhinav Gupta, and Laurens van der Maaten. 2018. Learning by asking questions. In *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 11–20. IEEE.

Karthik Narasimhan, Tejas Kulkarni, and Regina Barzilay. 2015. Language understanding for text-based games using deep reinforcement learning. In *Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing*, pages 1–11.

Andrew M Olney, Arthur C Graesser, and Natalie K Person. 2012. Question generation from concept maps. *Dialogue & Discourse*, 3(2):75–99.

Devi Parikh and Kristen Grauman. 2011. Interactively building a discriminative vocabulary of nameable attributes. In *Proceedings of the 2011 IEEE Conference on Computer Vision and Pattern Recognition*, pages 1681–1688. IEEE Computer Society.

Sudha Rao and Hal Daumé III. 2018. Learning to ask good questions: Ranking clarification questions using neural expected value of perfect information. In *Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, volume 1, pages 2737–2746.

Salvatore Romeo, Giovanni Da San Martino, Alberto Barrón-Cedeno, Alessandro Moschitti, Yonatan Belinkov, Wei-Ning Hsu, Yu Zhang, Mitra Mohtarami, and James Glass. 2016. Neural attention for learning to rank questions in community question answering. In *Proceedings of COLING 2016, the 26th International Conference on Computational Linguistics: Technical Papers*, pages 1734–1745.

G. A. Rummery and M. Niranjan. 1994. On-line Q-learning using connectionist systems. Technical Report TR 166, Cambridge University Engineering Department, Cambridge, England.

Burr Settles. 2012. Active learning. *Synthesis Lectures on Artificial Intelligence and Machine Learning*, 6(1):1–114.

Lanbo She and Joyce Chai. 2017. Interactive learning of grounded verb semantics towards human-robot communication. In *Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, volume 1, pages 1634–1644.

Behjat Siddiquie and Abhinav Gupta. 2010. Beyond active noun tagging: Modeling contextual interactions for multi-class active learning. In *2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, pages 2979–2986. IEEE.

Shashank Srivastava, Igor Labutov, and Tom Mitchell. 2017. Joint concept learning and semantic parsing from natural language explanations. In *Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing*, pages 1527–1536.

Shashank Srivastava, Igor Labutov, and Tom Mitchell. 2018. Zero-shot learning of classifiers from natural language quantification. In *Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 306–316. Association for Computational Linguistics.

Lucy Vanderwende. 2008. The importance of being important: Question generation. In *Proceedings of the 1st Workshop on the Question Generation Shared Task Evaluation Challenge, Arlington, VA*.

Sida I. Wang, Percy Liang, and Christopher D. Manning. 2016. Learning language games through interaction. In *Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics, ACL 2016, August 7-12, 2016, Berlin, Germany, Volume 1: Long Papers*.

Ludwig Wittgenstein. 1953. *Philosophical investigations. Philosophische Untersuchungen.* Macmillan.

Haichao Zhang, Haonan Yu, and Wei Xu. 2018. Interactive language acquisition with one-shot visual concept learning through a conversational game. In *Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 2609–2619. Association for Computational Linguistics.