Adversarial bandit for online interactive active learning of zero-shot spoken language understanding

Emmanuel Ferreira, Alexandre Reiffers-Masson, Bassam Jabaian and Fabrice Lefèvre
CERI-LIA, University of Avignon - France

Highlights

- **SoA** - Recent speech understanding systems rely on machine learning algorithms to train their models from large amount of data. Remaining difficulties: cost and time of data annotating and model porting to new tasks and languages
- **Novelty** - Zero-shot Semantic Parser, ZS learning method and semantic finite-state parser: combines an ontological description of the target domain and generic word embedding space for generalization
- **Current work** - Online adaptive process: refines initial model with policy learnt using an Adversarial Bandit algorithm

Zero-Shot Semantic Parser

Semantic Features Space (F) based on word-embedding

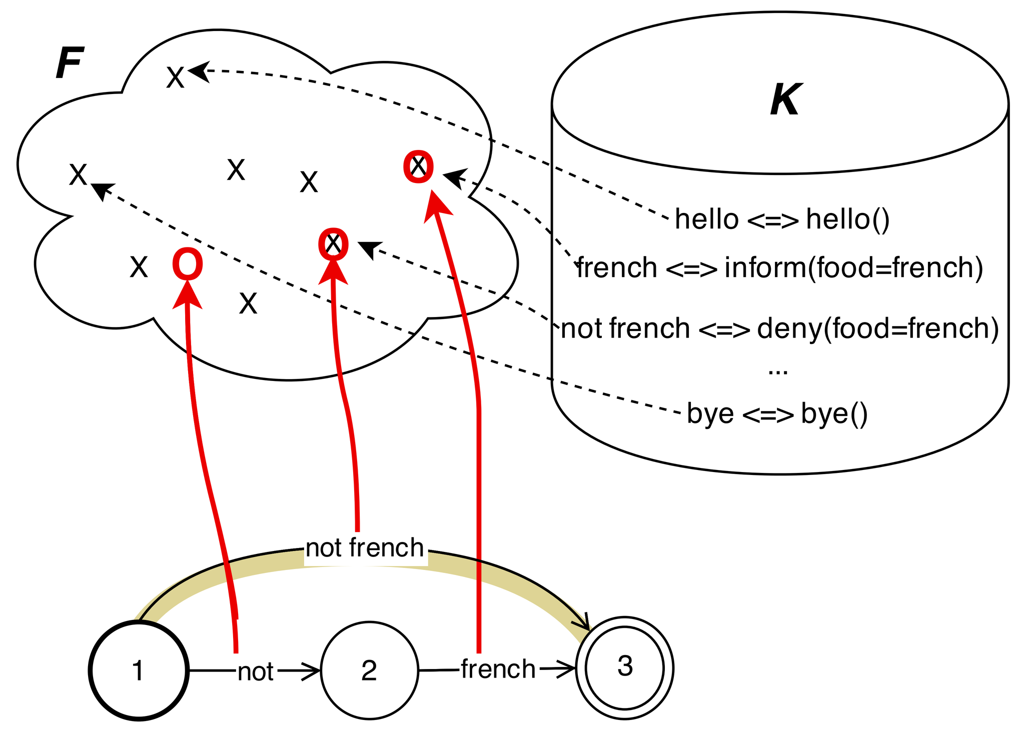
- Continuous representation of word learnt with neural network
  Sum operator used for word chunks
- Defines a metric space for generalization

Seed Semantic Knowledge (K)

- Domain-specific assignment table: task database + ontology of the domain
- Additional (reduced) dialogic knowledge

SLU Parsing

- K-NN classifier employed to attribute semantic hypotheses to every possible chunk of a test transcription
- Shortest-path estimated on the resulting chunks/semantic hypotheses graph



Online Interactive Refinement Problem

Action space

1. Skip: skip the refinement process.
2. YesNoQuestions: refine the model by considering yes/no user responses about the correctness of the detected DAs in the best semantic hypothesis.
3. AskAnnotation: ask the user to annotate the incoming utterance.

Loss function

$$l(i) := \underbrace{\gamma d'(i)}_{\text{system improvement}} + \underbrace{(1-\gamma)\frac{\phi(i)}{\phi_{max}}}_{\text{user effort}},$$

Extension to mixed strategies

$$\min_{\mathbf{p}\in\Delta(3)} E[l] = \sum_i p(i)l(i).$$

Adversarial Bandit environment

- System receives a user utterance and computes d_t;
- System chooses an action i_t, possibly with the help of external randomization;
- Once action i_t is performed, the system computes:

$\rightarrow$ Inefficiency measure $d'_t(i_t)$ with the collaboration of the user;
$\rightarrow$ User effort $\phi_t(i_t)$, which is the exchange count between the system and the user to compute i_t;
$\rightarrow$ Current loss is finally

$$l_t(i_t) = \gamma d'_t(i_t) + (1-\gamma)\phi_t(i_t).$$

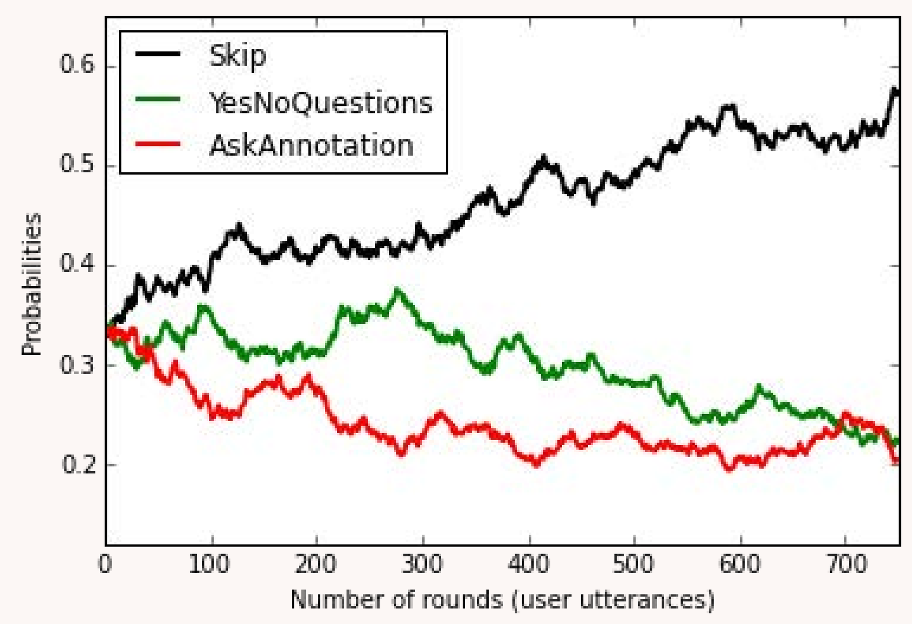
Goal: Find $i_1, i_2, \ldots$, such that for each T, the system minimizes the total loss:

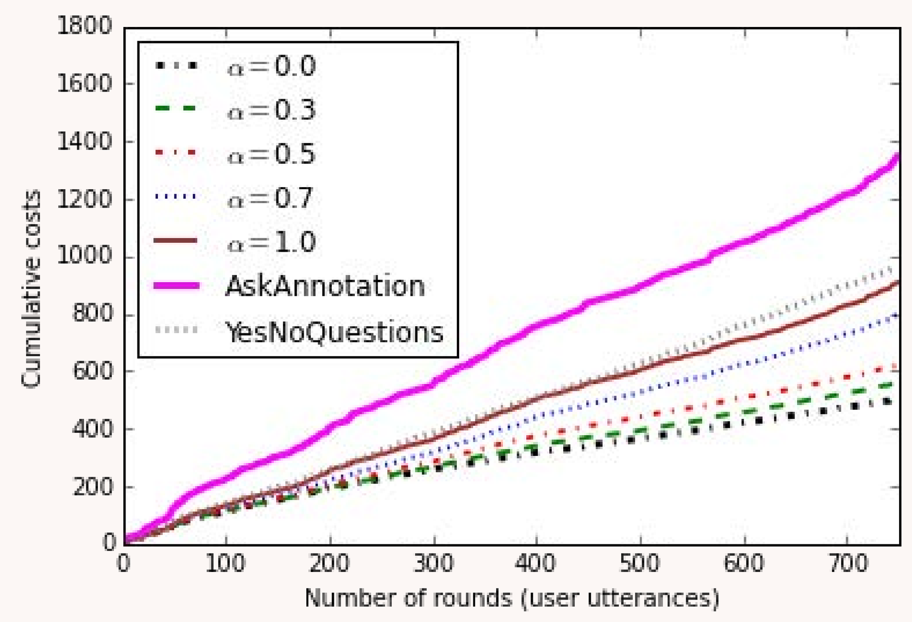
$$\sum_{t=1}^{T} l_t(i_t) = \gamma \sum_{t=1}^{T} d'_t(i_t) + (1-\gamma)\sum_{t=1}^{T}\phi_t(i_t).$$

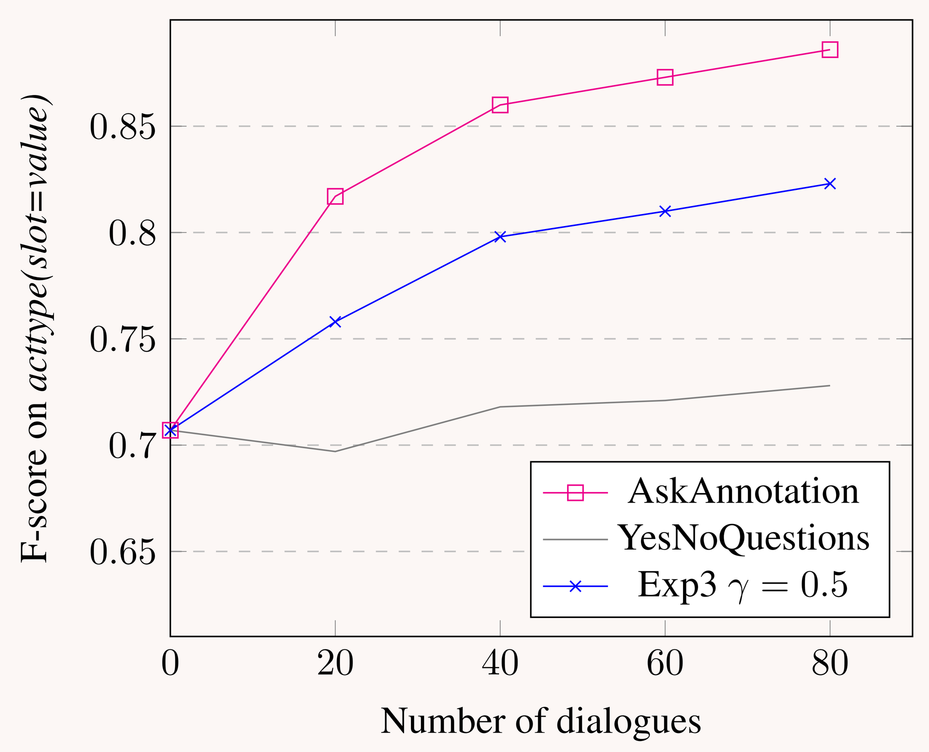
Solution: Randomized forecaster **Exp3**

Experimental Study on DSTC2

- **Ontology**: 16 dialogue act types, 8 slots and 215 values
- **Evaluation**: F-score performance on the test set (9890 user utterances)
- **Online Adaptation**: simulated from up to 740 transcribed training utterances







CONCLUSIONS

- Adversarial Bandit approach (and the use of the randomized forecaster **Exp3**) refines a zero-shot learning SLU, ZSSP
  $\rightarrow$ alleviate limited coverage of the domain specific semantics
- Efficient and practical way to formalise a trade-off between user supervision effort and system efficiency improvement
- Ongoing work:
  $\rightarrow$ integration in a live dialogue system with seed expert users
  $\rightarrow$ study effect over overall dialogue progress (task completion and user satisfaction)
  $\rightarrow$ relation with the dialogue manager strategy learning