Interactive Classification by Asking Informative Questions

Lili Yu, Howard Chen, Sida Wang, Tao Lei, Yoav Artzi
Intent Classification

• Classical classification problems operate on a single user input
• But natural language input can be underspecified and ambiguous
Intent Classification

- Classical classification problems operate on a single user input
- But natural language input can be underspecified and ambiguous
Intent Classification

• Classical classification problems operate on a single user input
• But natural language input can be underspecified and ambiguous

Goal: interact with the user to collect missing information
Challenges

- Interaction data is hard to get, often expensive

- Full-fledged dialogue modeling is data hungry and immature
Challenges

• Interaction data is hard to get, often expensive
  Can we bootstrap without user interaction data?

• Full-fledged dialogue modeling is data hungry and immature
  Can we design a lightweight model, that is constrained, but effective?
Interactive Classification

- Natural language intent classification
- We add binary or multi-choice clarification questions with predefined answer set
- At each turn, ask the most informative question, or return the best prediction

Templated Questions and Answer Set

- Do you have an online account? {Yes, No}
- Do you want to upgrade your service? {Yes, No}
- What is your phone operating system? {OS, Android, Windows}
## Intent Labels

<table>
<thead>
<tr>
<th>Question</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>What data limits apply when roaming internationally?</td>
<td></td>
</tr>
<tr>
<td>How do I purchase a High Speed Data Roaming?</td>
<td></td>
</tr>
<tr>
<td>Delete an individual message on your Samsung.</td>
<td></td>
</tr>
<tr>
<td>How do I sign up for Sprint Global Roaming?</td>
<td></td>
</tr>
<tr>
<td>Troubleshooting issues related to apps on iPhone.</td>
<td></td>
</tr>
</tbody>
</table>
Interactive Classification

Initial query $x$

Travel out of country.

Intent Labels

- What data limits apply when roaming internationally?
- How do I purchase a High Speed Data Roaming?
- Delete an individual message on your Samsung.
- How do I sign up for Sprint Global Roaming?
- Troubleshooting issues related to apps on iPhone.
Interactive Classification

Initial query $x$

Question $q^1$

Intent Labels

- What data limits apply when roaming internationally?
- How do I purchase a High Speed Data Roaming?
- Delete an individual message on your Samsung.
- How do I sign up for Sprint Global Roaming?
- Troubleshooting issues related to apps on iPhone.
Interactive Classification

Initial query $x$

Travel out of country.

Response $r^1$

Do you need to activate global roaming service?

Yes.

Question $q^1$

Intent Labels

- What data limits apply when roaming internationally?
- How do I purchase a High Speed Data Roaming?
- Delete an individual message on your Samsung.
- How do I sign up for Sprint Global Roaming?
- Troubleshooting issues related to apps on iPhone.
Initial query $x$:

Travel out of country.

Response $r^1$:

Do you need to activate global roaming service?

Yes.

Do you want high speed data roaming?

No.

Response $r^2$:

Got it! The article below might be helpful:

How do I sign up for Sprint Global Roaming?

Question $q^1$:

Do you need to activate global roaming service?

Question $q^2$:

Do you want high speed data roaming?

Intent Labels:

What data limits apply when roaming internationally?

How do I purchase a High Speed Data Roaming

Delete an individual message on your Samsung.

How do I sign up for Sprint Global Roaming?

Troubleshooting issues related to apps on iPhone.
Label Probability $p(y_i | X^t)$

**Simplifying assumptions**

1. User’s response depends only on the question asked and the underlying label
2. The model deterministically picks a clarification question given the interaction history
Label Probability

\[ p(y_i | X^t) \]

**Intent label**

**Interaction at time** \( t \)

**Simplifying assumptions**

1. User’s response depends only on the question asked and the underlying label
2. The model deterministically picks a clarification question given the interaction history

**Bayesian decomposition**

\[ p(y_i | X^t) \propto p(r^t | q^t, y_i, X^{t-1}) \cdot p(q^t | y_i, X^{t-1}) \cdot p(y_i | X^{t-1}) \]

\[ = p(y_i | x) \prod_{\tau=1}^{t} p(r^\tau | q^\tau, y_i) \]

**Initial label distribution**

**User response distribution**
Question Selection

Selection criterion
Select questions to maximize the interaction efficiency by maximizing the information gain

Intuitively: selecting the question that provides the most information about the intent label by observing its answer

Information gain computation
Can easily compute the information gain with
\[ p(y|x) \quad \text{and} \quad p(r|q,y) \]
Initial label distribution \quad User response distribution
Model the Distributions $p(y|x)$ and $p(r|q,y)$

- Model the distribution using text similarity
- Shared text embedding space
- Allowing to bootstrap for unseen questions, responses, and targets

Text piece

$$S(u, v) = \text{enc}(u)^T \text{enc}(v)$$

$$p(u|v) = h(u, v; \phi) = \frac{\exp(S(u, v))}{\sum_{u'} \exp(S(u', v))}$$

Concatenation of answer and question

$$p(y|x) = h(y, x; \phi)$$

$$p(r|q, y) = h(r\#q, y; \phi)$$

Treating each variable as text, not a categorical value
Model the Distributions $p(y|x)$ and $p(r|q,y)$

- Model the distribution using text similarity
- Shared text embedding space
- Allowing to bootstrap for unseen questions, responses, and targets

**Text Piece** and **RNN Text Encoder**

\[ S(u,v) = \text{enc}(u)^T \text{enc}(v) \]

\[ p(u|v) = h(u,v;\phi) = \frac{\exp(S(u,v))}{\sum_{u'} \exp(S(u',v))} \]

\[ p(y|x) = h(y,x;\phi) \]

\[ p(r|q,y) = h(r\#q,y;\phi) \]

Concatenation of answer and question

Treating each variable as text, not a categorical value
Policy Controller

Policy controller controls when and how to stop the interaction

Action space:

Ask an informative question

Stop interaction and return best label

Training:

Against user simulator; can extend to human-in-the-loop setting
Model Components

1. Label Probability
\[ p(y_i | X^t) \propto p(y_i | x) \prod_{\tau=1}^{t} p(r^\tau | q^\tau, y_i) \]

2. Question Selection
Maximize information gain

3. Model the Distribution
Model \( p(y|x) \) and \( p(r|q,y) \) using text similarity \( h(\cdot; \phi) \)

4. Policy Controller
Policy network \( f(\cdot; \theta) \), trained against a simulator to learn how to stop
Data Collection

• We crowdsourced data to bootstrap the learning process

• Two non-interactive tasks: initial query collection and question-answer pair collection

• We use the data to
  ✓ Train text embedding model
  ✓ Create a user simulator for training and evaluation
  ✓ Train a policy controller
**Domains**

**FAQ Suggestion**

**Scenario**
You would like to use your phone as hotspot for your laptop for some urgent work, but you are worried it gonna cost you lots of money.

**Model Predicted FAQ**
FAQs related to Wi-Fi hotspots on your samsung. Does it cost more to use my samsung as a mobile Wi-Fi hotspot?

- **How can I help you with your phone device?**
  1. app or feature information
  2. fee and charge
  3. troubleshoot device
  4. not applicable

- **Does you want to use phone as mobile wi-fi hotspot?**
  - yes

- **Here is the solution:** FAQs related to Wi-Fi hotspots on your samsung. Does it cost more to use my samsung as a mobile Wi-Fi hotspot?

**Bird Identification**

**Model Predicted Bird Type**
nashville warbler

**Scenario**
Here is the bird:
- little yellow bird with grey crown and black eyes

**Does the bird have yellow throat color?**
- yes

**Does the bird have yellow forehead color?**
- No

**What is the bird breast pattern?**
- solid
- spotted
- striped
- multi-colored

This bird is a: nashville warbler
Experiments

• Tasks: FAQ suggestion & Bird Identification

• Human evaluation: accuracy and user ratings

• Simulator evaluation: accuracy and cost analysis

• Two settings on FAQ: unseen labels + associated questions
  unseen labels (zero-shot)
Baselines

- No interaction:
  Directly return the best predicted intent label

- No initial query interaction:
  Interactions are not conditioned on the initial user query

- Ours (fixed turn):
  Stop asking questions after a fixed number of turns
Human Evaluation

- Improved accuracy > 90%
- Can generalize to unseen classes and utilize unseen questions
Human Evaluation

- Improved accuracy > 90%
- Can generalize to unseen classes and utilize unseen questions
- Text embedding improves accuracy
- Policy network effectively balances performance and effort
Our model receives higher ratings on **Naturalness** and **Rationality**.
Conclusion

✓ Interacting with the user to collect missing information
   By modeling user goal, user response, information gain, and termination policy

✓ Cheap: easy to bootstrap
   Non-interactive data collection, learning with simulator, zero-shot prediction

✓ Effective: adding interaction provides substantial improvement
   Demonstrated from human and simulator evaluation

Thank you!  https://github.com/asappresearch/interactive-classification